The implications of e-text resource development for Southern African literary studies in terms of analysis and methodology. Thesis to be supplemented with an e-text database and experimental CD-ROM.

By Graham Douglas James Stewart

Submitted in partial fulfilment of the requirements for the Degree of Doctor of Literature in Southern African Literature and Languages in the Centre for Southern African Literature and Languages, University of Durban-Westville, Durban.

February, 1999

Abstract

This study was aimed at investigating established electronic text and information projects and resources to inform the design and implementation of a South African electronic text resource. Literature was surveyed on a wide variety of electronic text projects and virtual libraries in the humanities, bibliographic databases, electronic encyclopaedias, literature webs, on-line learning, corcordancing and textual analysis, and computer application programs for searching and displaying electronic texts .The SALIT Web CD-ROM which is a supplementary outcome of the research - including the database, relational table structure, keyword search criteria, search screens, and hypertext linking of title entries to the electronic full-texts in the virtual library section - was based on this research. Other outcomes of the project include encoded electronic texts and an Internet web site.

The research was undertaken to investigate the benefits of designing and developing an e-text database (hypertext web) that could be used effectively as a learning/teaching and research resource in South African literary studies. The backbone of the resource would be an indexed "virtual library" containing electronic texts (books and other documents in digital form), conforming to international standards for interchange and for sharing with others. Working on the assumption that hypertext is an essentially democratic and anti-canonical environment where the learner/users are free to construct meaning for themselves, it seemed an ideal medium in which to conduct learning, teaching and research in South African literature.

By undertaking this project I hoped to start a process, based on international standards, that would provide a framework for a virtual library of South African literature, especially those works considered "marginal" or which had gone out of print, or were difficult to access for a variety of reasons. Internationally, the TEI (Text Encoding Initiative) and other, literature-based hypertext projects, promised the emergence of networked information resources that

could absorb and then share texts essential for contemporary South African literary research.

Investigation of the current status of on-line reference sources revealed that the digital frameworks underlying bibliographic databases, electronic encyclopaedias and literature webs are now very similar. Specially designed displays allow the SALIT Web to be used as a digital library, providing an opportunity to read books that may not be available from any other library. The on-line learning potential of the SALIT Web is extensive. Asynchronous Learning Network (ALN) programmes in use were assessed and found to offer a high degree of learner-tutor and learner-learner interaction.

The Text Analysis Computing Tools (TACT) program was used to investigate the possibility of detailed text analysis of the full texts included in the SALIT library on the CD-ROM. Features such as Keyword-in-context and word-frequency generators, offer valuable methods to automate the more time-consuming aspects of both thematic and formal text analysis.

In the light of current hypertext theory that emphasises hypertext's lack of fixity and closure, the SALIT Web can be seen to transfer authority from the author/teacher/librarian, to the user, by offering free access to information and so weakening the established power relations of education and access to education. The resource has the capacity to allow the user to examine previously unnoticed, but significant contradictions, inconsistencies and patterns and construct meaning from them. Yet the resource may still also contain interventions by the author/teacher consisting of pathways to promote the construction of meaning, but not dictate it.

A hypertext web resource harnesses the cheap and powerful benefits of Information Technology for the purpose of literary research, especially in the under-resourced area of

South African literary studies. By making a large amount of information readily available and easily accessible, it saves time and reduces frustration for both learners and teachers.

An electronic text resource provides users with a virtual library at their fingertips. Its resources can be standardised so that others can add to it, thus compounding the benefits over time. It can place scarce works (books, articles and papers) within easy access for student use. Students may then be able to use its resources for independent discovery, or via guided sets of exercises or assignments. Electronic texts break the tyranny of inadequate library resources, restricted access to rare documents and the unavailability of comprehensive bibliographical information in the area of South African literary studies.

The publication of the CD-ROM enables the launch of new, related projects, with the emphasis on building a collection of South African texts in all languages and in translation. Training in electronic text preparation, and Internet access to the resource will also be addressed to take these projects forward.

Declaration

This study represents original work by the author and has not been submitted in any form to another university. Where use is made of the work of others it has been duly acknowledged in the text.

Acknowledgements

I would like to thank the following people and institutions for their help and support:

My wife, Karin Stewart; my parents, Mary and Gordon Stewart; and my children, Peter, James and Nicola.

Johan van Wyk

Rashmi Jadhunundhan

Ian Sutherland

Antony Arkin

Joan Conolly

Kogi Naidoo

Wendy Gordon

Ansie Watkins

Theo Bothma

Nalinee Narain of the M.L. Sultan Research administration

The University of Durban-Westville Research Committee

The Head of Department and staff of the University of Essen Department of English

DAAD (German Academic Exchange Service)

Contents

Chapter 1 Introduction

Undertaking this study has involved some tension between my commitment to the demands of literary research and those of computing. What has kept me going is an overriding conviction that the output of this project will provide a useful tool for other researchers with an interest in South African literature, and entice a new generation of literary investigators into considering the past and present of South African literature as a rewarding area of study. I have been fortunate in finding the Centre for the Study of Southern African Literature and Languages (CSSALL), a literary research centre that shares this belief.

Researchers in British or American literature already have access to extensive text databases that they may utilise and may or may not overtly acknowledge in their work. Is it necessary to state explicitly that alongside your dog-eared Jane Austen novels you happen to have the electronic versions, with every word indexed in several ways and extensively cross referenced? Or that in your study of Dryden, you are able at the touch of a button to consult an electronic reference that provides you with every other occurrence of a word or phrase he has used by searching the text of all English poetry from 400 to 1900?

I share with other adherents of the "literary computing" faith the belief that information technology should be exploited by the humanities - and that it should be humanities scholars who take ownership of the process. This does not only entail a fight for a piece of real-estate in university IT budgets for the humanities although that is a necessary component of the mission. In literature studies we should be actively shaping the type of computing resources we require to achieve our research aims. That won't happen unless we teach ourselves the basic skills, and recognise clearly the benefits and the drawbacks of working in a virtual environment. The new bogey of "rationalised" humanities budgets can be turned to

our own advantage - computers offer a very real opportunity of doing more with less, the familiar slogan of the higher education cut-back commissars, and can support arguments for building IT resources in literary studies.

In short, the work on my project has constantly been subject to the sometimes conflicting pull of doing South African literary criticism, and doing it using computers. I have located my discussion of the design, structure and uses of a South African literature text database (referred to throughout as the SALIT Web, or simply SALIT) in the context of contemporary literary theory to show that computer-based research and learning offers not only a powerful information retrieval and analytic resource, but one that is uniquely suited to post-modern studies in literature. I have deliberately confined my own study to what is known as "legacy" information in computer terminology - i.e. the storage and retrieval of existing literary material - creative and critical works among others, but there is much evidence now of a new literary genre born of hypertext[1] that promises to be a truly unique form of expression - "tertiary orality" as it has been called in Orality studies.

Primary orality is typified by those groups who favour and the oral mode of communication, transmission of tradition, and record-keeping over the literate mode even when the latter is available to and commanded by them. Secondary orality is typified by the use of the electronic mass media - radio and television - which are, in turn, privileged over the written word. Tertiary orality, however, through the use of e-mail and the Internet, subjugates and transforms both the primary oral and the literate modes of communication to its peculiar ends. Hypertext can combine live on-line debate with its simultaneous "publication" in textual form. Literature teachers have experimented with the medium and their successes are discussed in Chapter 5.

---

[1] "electronic documents that present information that can be read by following many different connections, instead of just sequentially like reading a book." (Dowining 1997:181)

The SALIT Web is a hybrid literary resource that may be used in all the following ways:

- As a bibliographic database
- As a virtual library
- As a multimedia encyclopaedia
- As an archive for the preservation of marginalised texts
- As a textual analysis tool
- As a learning and teaching programme

To investigate my assumption that a computer-based electronic text resource would be a valuable learning, teaching and research tool, my study has involved a survey of the literature and computer programs dealing with all of these functions, together with the design and testing of programs appropriate to the needs of a range of learners and researchers in South African literature and literary history.

Two distinct categories of application can be discerned in the six uses listed above, viz: general access to information (bibliographic databases, virtual libraries, multimedia encyclopaedias and archives for the preservation of marginalised texts), and research and teaching methodologies (textual analysis tools and learning and teaching programmes). I have already alluded to a further intriguing category beyond the scope of my present study, but one that invites further investigation: the use of electronic text (in particular hypertext) for creative or imaginative composition - although, as we will see all hypertextual activity implies a creative engagement with the medium.

Because the possibilities of the SALIT Web are so extensive, I have limited my study to brief explorations of each of them. As I have produced working examples of most of the resources I discuss during the course of the study, this thesis constitutes only one aspect of

the project. This discussion represents one component of a multiple publication which includes the SALIT Web CD-ROM together with its database, retrieval interfaces, the Pringle full-text, the Internet access and the learning/teaching program. My hope is that this study provides at least enough insight into the potential of these resources for others in the South African literary field to use and adapt them for their own purposes.

The primary objective of this study is an investigation of the research and learning/teaching implications of a computer-based Southern African Literature research centre. At the heart of the project is a comprehensive textual database - the SALIT Web - providing several levels of information, starting with an on-line bibliographic catalogue, extending to encyclopaedia-like summary information (such as biographical sketches) and then to full texts of both primary and secondary material. Since the project coincides with the rapid development of the South African telecommunications infrastructure aimed at providing IT (information technology, i.e. computers, CD-ROMs and the internet) to the community at large, it is beginning to be possible to distribute literary resources to learners within institutions and out to the wider community. This is a context that favours an electronic rather than a print-based dissemination of literature. Although the entire spectrum of South African literature is now potentially available to all South Africans, there are formidable new obstacles to realising this ideal. In place of censorship we have impoverished libraries, rationalised school and university budgets and the rising cost of bound and printed books. While South Africans are building an open society, the printed word is threatened globally. "The displacement of the page by the screen is not yet total ..." writes Sven Birkerts, "but the large-scale tendency in that direction has to be obvious to anyone who looks" (Birkerts 1997: 3).

Around the world, electronic text databases are beginning to take a central place as resources for humanities scholarship, and what started as several disparate individual research projects in the 1980s has moved towards collaboration on universally accessible

formats like SGML (Standard Generalised Markup Language), so that primary and
secondary textual humanities data including books, manuscripts and journal articles can be
used by scholars across disciplines and languages. An electronic database can also store
graphic images and sound recordings, and be used as a platform for learning programmes.
Current practice and research in humanities computing are investigated in this study to
explore the implications for learning and research using the SALIT Web. Throughout the
discussion consideration is given to how the findings of humanities computing research has
influenced the design of the SALIT Web and how it may be used to underpin other
computer-based literary projects like the *South African Literary Encyclopaedia*, CD-ROM
publications of specific texts (e.g. Pringle's *African Sketches* and The Dube Web - see
Chapter 4) and networked learning programmes.

To this end, the following specific areas have been identified for investigation:

- A study of computer-based literary resources, including bibliographic databases,
  CD-ROM encyclopaedias and textual databases and their implications for the design
  and development of the SALIT database.

- A review of current practice in the representation of computer-based literary
  material including hypertext (on CD-ROMs and the internet) and encoding initiatives
  intended to promote the interchange and sharing of electronic resources.

- An investigation of the implications for literary scholarship and learning of the
  SALIT database including the digitising and structuring of texts and images for
  electronic publication and computer-based textual analysis.

- The determination of a methodology for the design and implementation of an
  Electronic Text Database in Southern African Literature (the SALIT database) in a

form compatible with future integrated computer networks in South Africa.

The SALIT database project implies two distinct (though related) activities. The first is a combination of reference and archival work: the collation of relevant commentaries biographical sketches, critical articles and archive materials across a wide range of texts. The collection then requires a coherent computer-based structure with extensive indexing and cross-referencing. Finally, a user interface must be designed to enable both novice and experienced users to find the information they want. The second activity is the exploration of the research possibilities inherent in the electronic literary text corpus itself: either as a tool for inter-textual literature study, or the analysis of a specific text to test text-critical assumptions.

The former process, building a structured and accessible "virtual library", is possible within the framework of the SALIT database (with the possibility of linking to related material in external sources - see Chapters 4 and 8). The latter process, primarily textual analysis, depends on the comprehensiveness of the material stored in the database. Ideally, the SALIT database could be hyperlinked to four levels or domains: (1) bibliographic information (2) encyclopaedia-like summary entries (3) full texts, graphics and sound-recordings, and (4) externally, to other electronic literary resources via the internet. In this study I will discuss each level separately for the sake of clarity, but from an electronic point of view, these would all fit seamlessly into the wider "docuverse"[2] now available via the internet.

An underlying assumption of the study is that the methodologies for literary research offered by encoded electronic text are congruent with post-modern methods of textual (and inter-textual) analysis. A distinctive feature of an electronic text resource is its capacity to

---

[2]   Nelson's term (1987) - see Page 16, 165 below.

be, simultaneously, a repository of documents and a method/means of investigation.

Foucault's "archaeological" and "genealogical" insights into the nature of warranted knowledge (Foucault 1972) provide a framework within which the electronic text resource can be examined as the basis for research methodologies that can "get behind the dominant or hegemonistic body of institutionalized and documented knowledge." (Hannabus 1996: 87-102). In his discussion of the ARTFL (the Project for American and French Research on the Treasury of the French Language) database - a large and comprehensive database of French language texts - Wolff (1994: 35-42) comments on the advantages for poststructural criticism of having ready access to a wide variety of texts. He cites an example of his own research into shifts in the meanings ascribed to the words "real" and "ideal" in a body of late nineteenth century literature. In doing so, he considers the role of computerised analysis:

> Perhaps one way to use computers effectively in textual analysis is to see how the text is able to manipulate how we read. Intertextuality from this angle would not be a static system of fixed signifiers but rather openings in the text that compel the reader to participate in the production of meaning. (Wolff 1994:39)

Wolff's comment identifies a characteristic of electronic texts - their ability to yield deconstructive perspectives - that makes them uniquely suited to provide solutions to Foucault's methodological problems for research into the "new history":

> These include: the building-up of coherent and homogeneous *corpora* of documents (open or closed, exhausted or inexhaustible corpora), the establishment of a principle of choice (according to whether one wishes to treat the documentation exhaustively, or adopt a sampling method as in statistics, or try to determine in advance which are the most representative elements); the definition of the level of analysis and of the relevant elements (in the material studied, one may extract numerical indications; references - explicit or not - to events, institutions, practices; the words used, with their grammatical rules and the semantic fields that they indicate, or again the formal structure of the propositions and the types of connexion that unite them); the specification of a method of

analysis (the quantitative treatment of data, the breaking-down of the material according
to a number of assignable features whose correlations are then studied; interpretative
decipherment, analysis of frequency and distribution); the delimitation of groups and sub-
groups that articulate the material (regions, periods, unitary processes); the determination
of relations that make it possible to characterize a group (these may be numerical or
logical relations; functional, causal, or analogical relations; or it may be the relation of the
"signifier" (*signifiant*) to the "signified" (*signifié*). (Foucault 1972: 10-11)

Computer-based analysis tools provide an efficient way of analysing texts to "investigate the
assumptions and underlying influences on the ways in which discourse embodies and shapes
meanings" (Hannabus 1996: 87). Studies of literature in which electronic texts have been
used in semiotic analysis to uncover discursive formations are reviewed here to evaluate
their appropriateness in a South African literary context, and in addition, some of the texts
in the SALIT Web Library collection are analysed to illustrate the capabilities of the textual
analysis tools (see Chapter 7).

Although the scope of a project aimed at producing a comprehensive database of Southern
African texts is clearly beyond the resources of any one research unit, or for that matter, any
one university or technikon, the importance of ensuring that endeavours in this area achieve
the goal of interchangeability of data cannot be over-emphasised. If scarce research
resources were to be used to produce a corpora of, say, late nineteenth century South
African poetry it would be nothing short of a tragedy to discover later that idiosyncratic
encoding of the texts prevented scholars from using one of the large international
collections such as the Oxford Text Archive to search for matching textual patterns in
contemporary European or Commonwealth poetry. Similarly, within South African
literature itself, it would be equally short-sighted to encode written and oral poetry in
mutually exclusive ways, precluding the extensive inter-textual analysis that electronic
encoding would make possible. Standardisation of encoding for electronic texts for research
is now a high priority for networked libraries, archives and data services.

This thesis argues the question of the relevance and applicability of methodologies arising from the use of electronic texts to literary research in South Africa, and whether the findings of scholars engaged in the use of full text databases for research purposes elsewhere has significance for the encoding of local texts.

It is argued further that the development of the SALIT Web would promote future research by allowing wider access both to published and previously unavailable texts, and by providing a comprehensive collection of Southern African texts on-line to researchers in community libraries, secondary education, universities and technikons. The database will be encoded (1) to facilitate author, title, date, event, biographical, bibliographic and keyword search and retrieve activities, and (2) to allow for the inclusion of full texts of some of the listed works, thus allowing for the close textual analysis. The coding would be sufficiently flexible to permit the future inclusion of additional biographical material, full-texts, graphics, sound and video recordings. To begin with, the SALIT Web would be developed on computers in the CSSALL (Centre for the Study of South African Literature and Languages) at the University of Durban-Westville and subsequently, as the corpus expanded, on a file-server that would allow for networking within the CSSALL and for users connecting to it via the Internet. At the same time, the database will provide the basis for the electronic version of the *South African Literary Encyclopaedia* to be published serially on a CD-ROM, containing full texts, graphics and sound recordings in addition to the content of the print-based *Encyclopaedia*. Ideally, the holdings of individual collections of Southern African Literature in libraries and museums such as NELM (The National English Literary Museum), NALN (Nasionale Afrikaanse Letterkundigemuseum en Navorsingsentrum), the Killie Campbell Library, the Don Africana Library and others could be made available via a network link, e.g. UNINET. The literature describes successful models of similar initiatives in the United States and Britain: the Center for Electronic Text in the Humanities (CETH) - Princeton and Rutgers Universities, Computing in the

Humanities and Social Sciences (CHASS) - University of Virginia, and the Oxford Text Archive (OTA) - Oxford University. In 1994, the Getty Art History Information Program discovered, during the drafting of a white paper entitled "Humanities and Arts on the Information Highways: A Profile", that little was known about the range of humanities projects exploiting information technologies. Their current group of projects under the title of the "Networked Access Project" (Bearman 1995) is likely to be highly influential in formulating a research agenda for humanities computing in the United States, and as their work is so pertinent to the SALIT project, a brief description is included here.

In an attempt to formulate a research agenda for humanities computing, Bearman co-ordinated the project within which eight researchers were commissioned to evaluate current access to electronic resources in the arts and humanities. Bearman suggested the evaluation be done "in a sequence from the beginning of a research process through to the archival life of the product" (Bearman 1995: 2).

The steps that emerged were identified as: **Tools** - tools for the creation of scholarly work; **Representation** - methods of representing knowledge; **Conversion** - issues in converting traditional formats to digital form; **Discovery** - information discovery and retrieval; **Image** - multimedia representation and retrieval; **Learning** - teaching and learning; **Socio-economic** - new social structures for supporting digital scholarly activity; and **Archiving** - archiving of digital records. Not all these areas are of direct relevance to this study, but three of them - **Tools and Representation**, **Discovery**, and **Learning** - have application to the SALIT database, and correspond to the six facets of the SALIT Web: bibliographic database, virtual library, multimedia encyclopaedia, archive for the preservation of marginalised texts, textual analysis tool, and a learning and teaching programme. As in the SALIT project, Bearman's facets or areas imply discrete categories, but each is closely interwoven at several levels, and all underpin humanities, and specifically, literary scholarship.

As there is no "macro" level project in South Africa on the scale of the Getty project, the CSSALL's SALIT database has the potential to play a significant role in guiding future developments in the building of a corpus of South African electronic texts. The *Encyclopaedia* project, which shares some of the content material, is highly participatory in nature, drawing on the expertise of literature and language scholars from a wide range of South African academic institutions. While pursuing the primary goal of producing the country's first multicultural literary reference work, it could also be a vehicle for the sharing of information and skills in the production of electronic texts. The *Encyclopaedia* project is led by Prof. A.J. van Wyk, Dr J-P Wade and Dr J. Smit (with Prof. D. Killam, University of Guelph, Canada as project advisor) and the Department of English at the University of Essen, Germany. Launched in 1995 following a colloquium, subsequently published as *Rethinking South African Literary History* (Smit, Van Wyk & Wade 1996), publishers and academics from several South African university Language and Literature departments subsequently participated in a workshop in June 1996. It was decided that van Wyk's original document *Concise Historical Survey: South African Literature* (1996b) would form the basis of the new *Encyclopaedia* and that the editors could proceed with the identification of project leaders for the various encyclopaedia sections, and allocate entries to a large number of individual contributors.

Conceived originally as a convenient management device to help organise the *Encyclopaedia*, the SALIT database has now become a separate resource in its own right (see Chapter 10). This is a phenomenon peculiar to the development of other electronic resources as the discussion of on-line bibliographic databases and encyclopaedias in Chapters 2 and 3 will show. While the development of the SALIT Web proceeded parallel to the main print-based project using identical content, it displayed significant additional properties including powerful search, retrieval and cross-referencing features, hypertext links, the capacity to house a full text library, photographs and sound recordings. Although the SALIT Web CD-ROM will eventually be published using the international text encoding

and exchange "meta-language" Standard Generalised Markup Language (SGML - see Chapter 6), it was necessary to design and create the SALIT database to structure and manage the information. The discussion in Chapter 3 addresses the distinctions (and similarities) between databases and encyclopaedias pointing out that essentially, an on-line (or CD-ROM) encyclopaedia *is* a database containing specific reference and summary information structured according to the conventions of encyclopaedia form.

The relationship between databases and SGML has such critical significance for the SALIT database project that a brief explanation is called for here (this issue is further elaborated in Chapter 6 and Chapter 10). To begin with, in electronic text terms, a database is old technology while SGML encoded text is new. But there is more to this distinction than mere novelty. While databases are designed as highly versatile "index card" systems, fitting information into pre-determined categories, SGML puts the text at the centre, encoding it in such a way that all the powerful indexing and cross-referencing capabilities of a database are retained. In SGML every word is potentially a "field" or a "key term". Moreover, in the TEI (Text Encoding Initiative) "dialect" of SGML (see Chapter 6), the coding has been painstakingly developed to tag characteristics unique to literary texts. Combined with a hypertext system, the encoded text has the potential to be used in a variety of ways: for detailed analysis, linking to explanatory material or for cross-referencing almost any other networked information source. TEI encoded electronic texts will ultimately the basis of the SALIT Web Library collection, with distinguished publications such as the Electronic *Oxford English Dictionary* CD-ROM and the Cambridge University Press *The World Shakespeare Bibliography* (Harner 1996) as precedents. Above all, SGML encoding ensures universal exchangeability of data, independent of proprietary computer software packages.

Underpinning the entire SALIT project is the new computer-based non-linear writing form known as hypertext. The theoretical background to hypertext has been explored by Landow

(1994) and by Kolb (1994) and its applicability to postmodern literary and critical theory chimes with the reconception of the relations between author and writer articulated by Barthes (1997). Hypertext theory and postmodern theory "both grow out of dissatisfaction with the related phenomena of the printed book and hierarchical thought" (Landow 1994:1) . David Kolb argues that a hypertext web "can in its way be a strong symbol of the unlimited fecundity of signs, of the varied individuation of texts and entities, and of the endlessness of redescription and reincorporation" (Kolb 1994: 337).

What is hypertext? Dowining provides the following definition: "electronic documents that present information that can be read by following many different connections, instead of just sequentially like reading a book" (Dowining 1997: 181). Hypertext is a crucially important feature of the SALIT Web and its integrated the electronic full-text publications, both as a theoretical framework and as the paradigm for its design. So before proceeding to an outline of the thesis, I describe some of its key features and briefly trace its history.

To start with, it may be helpful to use a few analogies to expand the definition given above, in order to dispel the notion that hypertext *as a process* is in any way alien to literary scholarship. For instance, a library may be thought of as a type of manual hypertext - a user may browse through titles on the shelves and open and read all or portions of some of the texts themselves, propelled by the "hypertext of the mind" (Rockwell 1996: 7) that informs scholarly investigation. A similar activity that is "hypertextual" and yet in no way electronic, is a reader's reference to footnotes and endnotes within a text. For the reader, marginal information like this serves to amplify and contextualise the main body of the text. Other familar metatextual features are indices and contents pages. Electronic hypertext accelerates the familiar process of mental association between links in the body text and references "outside" by instantly displaying the reference to the reader (typically, in a "window").

Although the process may be familiar, hypertext brings a significantly more powerful

element into literary investigation. As the definition implies, it is the *connections* that the reader may follow that characterises hypertext - a reader may access not only references in the text, but the *full texts* of those references themselves, so encouraging a tangential in-depth reading of the new text. It can be seen then that the library analogy is closer to the experience of using hypertext than that of the footnotes, although both comparisons are valid.

Further examples of texts that pre-date hypertext and yet share some of its characteristics are reference works like encyclopaedias that contain articles with links (the "see also"-type of references) to related material elsewhere in the text. The structures of print-based and hypertext encyclopaedias are explored more fully in Chapter 3.

Rockwell (1996: 10) suggests the following table contrasting text with hypertext:

Text	Hypertext
Sequential	Non-sequential (browsing, navigating, "surfing")
Paper	Computer
Read	Browse or navigate
Table of Contents	Map
Text and images	Text, images, audio, video and interactive procedures
Easy to carry and use	Needs an expensive computer and takes time to install, launch and use.

As can be seen, Rockwell implies a number of criticisms of hypertext, including its tendency

to discourage full readings (although not necessarily inhibiting close reading of selected passages) and its relative inaccessibility compared to the book. Most commentators emphasise the non-sequential nature of hypertext as its most important distinguishing feature: "Hypertext is *non-sequential*; there is no single order that determines the sequence in which the text is to be read" (Nielsen 1995: 1).

Welsch (1992: 615) quotes Nelson's 1987 definition of hypertext as *nonsequential writing -* "text that branches and allows choices to be made, [and] is best read at an interactive screen" (Nelson 1987: 2). The word "hypertext" can be traced back to the 1960s and was coined either by Nelson or by Engelbart, another early pioneer of electronic text systems. The first writer to have proposed what we would now describe as a hypertext system is Vannebar Bush who postulated a device that would be an "enlarged intimate supplement" to our memory called "Memex"(Bush 1945). Although Bush's "Memex" is a mechanical information device, his prediction that it could mimic "the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain" coincides with the design of present day hypertext systems, including the Internet. Bush anticipated with remarkable accuracy the type of environment used, for example, in my own development of the Folio Views "front-end" to Pringle's *African Sketches* as part of this project:

> A special button transfers him immediately to the first page of the index. Any given book
> of his library can thus be called up and consulted with far greater facility than if it were
> taken from a shelf. As he has several projection positions, he can leave one item
> in position while he calls up another. He can add marginal notes and comments, [...] just
> as though he had the physical page before him. (Bush 1945:14)

In 1965 Ted Nelson in *Literary Machines* ([1965] 1987) described his vision of a major hypertext resource called Project Xanadu. Although it has never been implemented, its similarity to current hypertext systems is clear: he proposed a browser/server architecture where people would pay to connect to the Xanadu site where the information was stored, a

site that would constitute a single "docuverse" where everyone had access to all published information at the same time. Xanadu also included "transclusion", a form of link where one document could have part of another document embedded within it, anticipating the accounting systems now operating on certain Internet sites that bill the "client" for the time used or amount of information downloaded.

Nelson's cover notes from *Literary Machines* conveys the fervent tone of his vision: "... Project Xanadu, an initiative toward an instantaneous electronic literature; the most audacious and specific plan for knowledge, freedom and a better world yet to come out of computerdom; the original (and perhaps the ultimate) Hypertext System." (Nelson [1965] 1987). Since then, the development of hypermedia systems (Intermedia at Brown University and the Guide document Browser) in the 1980s and the Internet in the 90s have, together with the rapid pace of computer hardware improvements brought these hitherto speculative resources into everyday use.

Robert Coover's definition of hypertext in his 1992 essay "The end of books" (Coover 1992) is another frequently cited explication of the new genre:

> "Hypertext" is not a system but a generic term, coined a quarter of a century ago by a computer populist named Ted Nelson to describe the writing done in the nonlinear or nonsequential space made possible by the computer. Moreover, unlike print text, hypertext provides multiple paths between text segments, now often called "lexias" in a borrowing from the pre-hypertextual but prescient Roland Barthes. With its webs of linked lexias, its networks of alternate routes (as opposed to print's unidirectional page-turning) hypertext presents a radically divergent technology, interactive and polyvocal, favoring a plurality of discourses over definitive utterance and freeing the reader from domination by the author. Hypertext reader and writer are said to become co-learners or co-writers, as it were, fellow travelers in the mapping and remapping of textual (and visual, kinetic, and aural) components, not all of which are provided by what used to be called the author. (Coover 1992).

The terms "hypertext" and "hypermedia" have thus far been used interchangeably, but a

distinction can be made if one thinks of "Hypermedia" as "hypertext with multimedia added". Multimedia includes graphics, sound recordings, videos and application programs that may also form links in the "nonsequential" information environment that constitutes an electronic text. Nielsen explains the term thus:

> Since many of the current systems actually also include the possibility for working with graphics and various other media, some prefer using the term *hypermedia*, to stress the multimedia aspects of their system. (Nielsen 1995: 5)

The SALIT database should be viewed in the light of current hypertext theory that emphasises its lack of fixity and closure. The open, democratic nature of hypertext transfers authority from the author/teacher/librarian to the user by offering free access to information and so weakens the established power relations of education and access to education. As hypertext consists of independent units of information ("lexias") that co-exisit without any one being privileged over another, navigation amongst them lies in the hands of the user. This characteristic may be compared to the notion founded in post-structural theory that apparently closed structures such as narratives operate within a larger field that they do not control. Users of hypertext may construct their own meanings from the information units contained in it, so the material in the SALIT resource, for example, is essentially anti-canonical and each text is surrounded by a "fluid discourse" of commentaries, reviews, and annotations which the user may select, add to or ignore. The resource has the capacity to allow the user to examine previously unnoticed, but significant contradictions, inconsistencies and patterns and construct meaning from them. Yet the resource may also contain interventions by the author/teacher consisting of pathways (keyword links, for example) that may *promote* the construction of meaning, but not *dictate* it.

As an Information Technology, hypertext (and by association the SALIT Web) requires critical attention as a powerful formation in itself. Lyotard (1979) has recoiled from the notion that the new information technologies appear to confine the text to a pre-

programmed set of possibilities, controlled by a programmer. Hypertext, however, resists the authoritarian potential of the technology by being an incorrigibly open system, its potential for independence evidenced by the democratic (and chaotic) nature of the Internet which is a hypertext resource on a global scale. By asking critical questions like:

> How is it that we actually use the technology?
> What language do we use?
> How do we interact with one another, and the technology?
> What impact does this have on us - on all of us?
> (Wilson 1997: 25)

we can contextualise our use of the SALIT Web within the area of Information Technology, and reflect on the extent to which it matches our aims of wider access, learner-centred teaching and the nature of the interaction between the user and the resource.

The term "hypertext" is used throughout this study to refer also to what Nelson (1989) calls "hypermedia" in a later extension of his original definition. The SALIT Web also includes non-text resources including "text, graphics, audio and video [that] can now come alive in unified, responding, explorable new works that present facts and ideas: hypermedia" (Nelson 1989: 3-5). Although the SALIT Web could be regarded as "hypermedia" according to this definition, hypertext is more generally used in the literature, and seemed to be a more appropriate choice in identifying the substance of this project.

The following timeline highlights the most important developments in the history of hypertext:

A Hypermedia Timeline	
1945	Vannevar Bush (The Science Advisor to President Roosevelt during World War II) proposes MEMEX, a conceptual machine that can store vast amounts of information, in which users have the ability to create information trails, links of related texts and illustrations, which can be stored and used for future reference.
1965	Ted Nelson coins the word "hypertext"
1967/8	Andy van Dam and others at Brown University build two of the first hypertext systems: the Hypertext Editing System and FRESS. (http://twine.stg.brown.edu/projects/hypertext/landow/HTatBrown / BrownHT.html)
1968	Doug Engelbart demonstrates NLS, a hypertext system.
1975	ZOG (now KMS), a distributed hypermedia system, debuts at Carnegie-Mellon.
1978	The Aspen Movie Map, the first hypermedia videodisc, demonstrated by Any Lippman of MIT's Architecture Machine Group. The videodisc created an environment in which the user can drive around the town of Aspen.
1981	Ted Nelson conceptualizes "Xanadu", a central, pay-per-document hypertext database encompassing all written information. Read the Xanadu FAQ at http://jolt.mpx.com.au:70/0h/faq.html.
1984	Telos introduces Filevision, a hypermedia database for the Macintosh.
1985	Janet Walker creates the Symbolics Document Examiner.
1985	Intermedia, a hypermedia system, is conceived at Brown University by Norman Meyrowitz and others. Intermedia was used by Landow to create instructional hypertexts for literature courses. (See http://twine.stg.brown.edu/projects/hypertext/landow/HTatBrown/ BrownHT.html)
1986	OWL introduces GUIDE, a hypermedia document browser.

A Hypermedia Timeline	
1987	Apple Computers introduces HyperCard, the first widely available personal hypermedia authoring system. HyperCard was bundled with every new Apple Macintosh computer so putting an easy to use hypermedia environment into the hands of many instructors and researchers.
1987	The Hypertext '87 Workshop is held in North Carolina.
1989	Autodesk, a major CAD software manufacturer, takes on Xanadu as a project.
1989	Tim Berners-Lee proposes the World-Wide Web project. (See the proposal: http://www.w3.org/hypertext/WWW/Proposal.html) "HyperText is a way to link and access information of various kinds as a web of nodes in which the user can browse at will. It provides a single user-interface to large classes of information (reports, notes, data-bases, computer documentation and on-line help). We propose a simple scheme incorporating servers already available at CERN." (The opening paragraph of the proposal available at the URL above)
1990	ECHT (European Conference on Hypertext).
1992	Autodesk drops the Xanadu project.
1993	Print Encyclopaedias are surpassed in sales by Hypermedia Encyclopaedias.
1993	*A Hard Day's Night* from Voyager, becomes the first digital video hypermedia work to be published and distributed via compact disc.
March 1993	Hyperfiction. Hyperfiction is on the cover of the *New York Times Book Review* (29 August, 1993) "Hyperfiction is a new narrative art form, readable only on the computer and made possible by the developing technology of hypertext and hypermedia. Not all adults have familiarized themselves with hypertext, but most children have, for it is the basis of many of their computer games and is fast becoming the dominant pedagogical tool for our digitalized times." (p. 8)
April 1993	International Workshop on Hypermedia and Hypertext Standards, Amsterdam.

A Hypermedia Timeline	
June 1993	NCSA Mosaic 1.0 for X Windows released by the National Center for Supercomputing Applications.
August 1993	First World-Wide Web developers' conference in Cambridge, Massachusetts.
November 1993	Hypertext Conference in Seattle, Washington. Ted Nelson speaks as the guest of honour.
March 1994	World-Wide Web byte traffic surpasses Gopher traffic on the NSFnet.
May 1994	First International World-Wide Web Conference in Geneva.
	Jim Clark and Marc Andreessen form Mosaic Communications Corporation.
June 1994	World Conference on Educational Multimedia and Hypermedia in Vancouver, Canada.For information email aace@virginia.edu.
September 1994	European Conference on Hypermedia Technology in Edinburgh, Scotland.For information email echt94@inesc.pt.

(Rockwell 1996: 13-15)

Explanation of some of the more common technical terms used in hypertext and hypermedia can be found in the Glossary of Terms (Appendix F), but a brief review of the underlying structure of hypermedia and the terminology required to understand its operation is presented here. A hypermedia is composed of nodes (or documents) of information that are linked. Each node (or document) can have one or more links entering or leaving it. The network of linked nodes can be represented as a web. Using a browser (e.g Folio Views for the Pringle *African Sketches* document and the *Encyclopaedia*; Netscape Navigator for Internet access to the SALIT Web), the reader may take a route through the information (from one node to the other). Routes (or trails) of interest can be marked and replayed using the "history" feature. A simple link has at least one anchor point which is the information from which the link is launched (usually by clicking the mouse button); it will also have a

"destination". A destination is the target of the link and must be encoded in a way that allows the computer to traverse the link when it is activated. In HTML[3] a link is encoded by bracketing the text to be the anchor with the <A> code and providing the destination (the HREF) in the anchor-on code":

<A HREF= "http://www.udw.ac.za:80/~stewartg/alternat.html" *Alter*nation Journal </A>

An HTML browser (e.g. Netscape) hides the codes from the reader, and will typically underline the anchor text in blue. If the reader clicks the mouse button on the anchor text the HTML browser links to the target page and loads it in the window. The above link in the CSSALL Internet Home Page (see Chapter 9 and Chapter 10) takes the user to the *Alternation* journal contents pages. The same process and terminology applies in the case of non-HTML browsers like Folio Views, used for part of this project.

To maximise the potential of the medium for learning/teaching and research, the SALIT resource aims to use the medium of hypertext to promote user accessibility at many different levels (novice, student, teacher, researcher) and allow the user to annotate and customise the contents. To be a valid scholarly resource, it must also conform to conventional publication standards and assure the user of texts of known provenance. Guides and commentaries should also conform to the highest academic standards and be clearly signposted as "authorial interventions".

At the beginning of this chapter, I categorised the uses of the SALIT into two areas: (1) general access to information and (2) research and learning resources. As now may be clear from the foregoing discussion, non-linearity is a fundamental characteristic of the

---

[3] HTML Hypertext Markup Language - the text encoding system used for Internet documents.

SALIT resource, so there remains a substantial overlap amongst the five sub-sections that constitute (a) and (b). Nevertheless, I have imposed a sequential ordering based on the chronological development of Electronic Information Sources (Chapters 2 - 4) and the fairly distinctive areas of learning and literary research in Analytical and Learning/Teaching Tools (Chapters 5 - 7). These broad sections culminate in an account of the structure and uses of the SALIT Web (Chapters 8 - 9). A third major division - the SALIT Development (Chapter 10) - documents the process of designing and developing the SALIT CD-ROM, and is supplemented by further appendices on various aspects of the production of electronic texts. Also included is a copy of a conference paper that concentrates on the "virtual library" capabilities of the SALIT Web, and a comprehensive glossary of terms.

The following table outlines the structure and summarises the content of the thesis:

Area	Chapter	Content
	1	Introduction
General access to information	2	Bibliographic databases
	3	Encyclopaedias
	4	Literature webs
Analytical and learning/ teaching tools	5	Learning
	6	Encoding
	7	Concordancing & textual analysis
	8	Using the SALIT Web
	9	The way forward: Internet access
	10	Development of the SALIT Web
	11	Conclusion

Area	Chapter	Content
Appendices: SALIT development	Appendix A	Notes on installing the Beta 01 Version of the *SALIT Web* CD-ROM
	Appendix B	Conversion of the database from linear to relational form
	Appendix C	Producing electronic texts: A practical course outline
	Appendix D	"A Home for our Marginalised Past" Paper presented a the South African English Academy Conference, 1998.
	Appendix E	Evaluation of application programmes for displaying and searching electronic text
Glossary of terms		
References		

Chapters 2 to 4 (General Access to Information) deal with various on-line literary resources like bibliographic databases and CD-ROM encyclopaedias as they relate to the development of the SALIT database. The structuring of on-line literary material is discussed, as well as search and retrieval characteristics that have shaped the design of the present project.

Chapters 5 to 7 (Analytical and Learning and Teaching Tools) take the discussion further, examining the SALIT database as a platform for learning and for linking resources in new ways. Chapter 6 reviews the encoding of texts for literary research and emphasises the key role of the TEI (Text Encoding Initiative) in providing a model for future electronic text development and publishing, while computer-based textual analysis methodology is examined in Chapter 7, using electronic version of Pringle's *African Sketches* (1834) and Mda's *The Hill* (1995), scanned, digitised and encoded specifically for this project as examples.

Chapters 8 and 9 provide an insight into the design and uses of the SALIT Web and its

potential for future development. Chapter 10 traces the technical development of the SALIT Web. The characteristics of databases are reviewed, and the various stages of the development of the SALIT database are documented. Some conclusions are drawn in Chapter 11.

Appendix A provides a guide to the installation of the experimental Beta 01 version of the CD-ROM. Appendix B details the conversion of the linear database to its final stage as a relational database while Appendix C presents the outline of a course developed for the preparation of electronic texts and mentions collaborative projects between the CSSALL and M.L. Sultan Technikon. "A home for our marginalised past: Creating a virtual library of South African literature" is included as Appendix D. This is a copy of a paper that addresses the SALIT Web's "virtual library" component, which I delivered at the South African English Academy Conference, "English at the turn of the Millennium" on 16 September 1998. Appendix E contains evaluations of the software applications used during the course of the project. This is followed by a comprehensive glossary of terms used in the thesis.

Chapter 2 Bibliographic Databases

Until the 1990s, most researchers were the passive recipients of bibliographic information provided for them by reference librarians who acted as intermediaries between them and bibliographic databases. The Internet has now profoundly changed this relationship, by propelling researchers into active users of on-line bibliographic resources. "Searching" and "retrieving", once the exclusive realm of the librarian has become commonplace to internet users, and search engines like Yahoo and AltaVista are now familiar tools for finding information on the Net. The advantages of this sudden leap forward are obvious, not least because the researcher is now free to access information directly, from his or her desktop, and at low cost. The recent progress made in "user friendly" database technologies has a direct bearing on the SALIT Web project described in this thesis. Without the existence of development tools that can be used by non-technicians, the present project would have been beyond the scope of most humanities scholars ("literary computing" as Mark Olsen has pointed out tended to be a marginal activity in literary scholarship - Olsen, 1993). There is still, however, a significant gap between the vast, disorderly "virtual library" of the internet and the organised, authoritative world of the bibliographic database. Even using the internet to tap into the on-line public access catalogues (OPACs) of South African libraries brings the user back to the daunting DOS-based menu-driven terminals we have grown accustomed to in libraries since the demise of the card catalogue.

The strength of the SALIT Web lies in its hybrid nature: using new network-related computer technologies, it is now possible and even desirable, that a literary resource should combine the search and retrieval power similar to those of a bibliographic database but enhance this with multimedia and communication elements. Photographs, maps, animations and video material have been linked to the SALIT Web bibliography, and its interactive

learning potential is explored further in Chapter 5. The advantages of integrated projects can be seen in the success of The Library of Congress American Memory archive, where digitised resources (photographs, images of paintings) and bibliographic access have been merged to provide access flexible enough to meet the needs of users from secondary school to university level. One of the Library of Congress interactive on-line sites is also illustrated in Chapter 5. Similarly, the SALIT Web includes many of the features of a bibliographic database such as author and title details, but also contains additional information such as biographies and photographs together with links to full-text electronic versions of the books, reviews and other documents located in the Library section. A discussion of the elements of a "conventional" bibliographic database will serve to highlight the unique characteristics of the SALIT Web while emphasising some very fundamental differences.

Before considering the structure and uses of the most comprehensive bibliographic database of South African literature, the NISC *South African Studies* CD-ROM (1996), it may be useful to review the history of the bibliographic database in more general terms. A striking feature of this history is that the first computerised bibliographies in the 1960s, *Chemical abstracts* and *Index medicus*, were not originally developed as usable databases as we now know them. The computerised records were at first used exclusively to assist the *print* publication of the bibliographies - the digital database was regarded merely a means to this end. The fact that the library world recognised only later that the computerised database was a new type of *publication in itself* with information searching capabilities far beyond its print-based counterpart, matches the possible confusion of some literary scholars about the nature of a resource like the SALIT Web. As with the first bibliographic databases, it takes time for users to adjust to the potential of a new and unique medium. For most humanities scholars, databases (like word processed files) are still just a means to an end - print-based publication. While the present project may indeed lead to some conventional print publications, e.g. *The South African Literary Encyclopaedia* and previously unpublished translations, the database resource itself, accessed via the computer, has a value and a usefulness intrinsic to itself.

In the case of the development of bibliographic databases, it took less than ten years for a new intermediary industry to be established - the database vendor - who obtained the bibliographies from the publishers and made these available to libraries and research institutions via telecommunications channels. By the 1980s, the invention of the CD-ROM, an "optical disk" storage medium that uses laser technology to store information, had made the distribution of databases *in their entirety* quite feasible. Bibliographic databases had moved far beyond being merely the scaffolding for print publications - they had largely superceded them, and a new more versatile medium had been born. The SALIT Web can also be seen as an intermediary, pulling together comprehensive information on South African literature for a new type of end-user: a computer-literate scholar or researcher accustomed to immediate Internet-like access to full-text and multimedia material. As discussed in the next chapter, the SALIT Web is more like an encyclopaedia than a bibliography.

To return to our discussion of conventional bibliographic databases: print-based bibliographies are updated through the distribution of separate supplements that are added on (or "cumulated") over time while the CD-ROM versions have sufficient storage capacity to include the full updated database. The *South African Studies* database is updated every three months, and the new CD-ROM distributed to subscribers who then return their copies of the previous disk to the publisher. What is the *South African Studies* CD-ROM bibliography? Published for the first time in 1996 by the National Enquiry Services Centre (NISC) in South Africa, *South African Studies* consists of 11 South African databases including the South African National Bibliography (SANB), the databases of the National English Literary Museum (NELM) and *Knipkat*, an index to the press cuttings service of the Nasionale Afrikaanse Letterkundige Museum en Navorsingsentrum (NALN). The currency of the bibliographies is ensured by regular three-monthly updates and the annual subscription for single-user access is about R4000. This "anthology" CD-ROM has the advantage of including a number of valuable South African literary bibliographies on one

disk with a common user interface (ROM-WRIGHT) and this inclusiveness matches some of the intentions of the SALIT Web - multiple information sources "under one roof". However, the nature of the information provided by the *South African Studies* CD-ROM is chiefly bibliographical: indices of sources available, and does not include material beyond this level as is the case with *the SALIT Web*, e.g. biographical sketches, full-texts, photographs, etc. The value of the *South African Studies* CD-ROM lies in the comprehensiveness of the bibliographic entries and their congruity with national and international bibliographic practice. Each entry, by virtue of the standard practice of their source suppliers (e.g. the SANB), is encoded in accordance with the aims of Universal Bibliographic Control (UBC) promoted by the United Nations Educational, Scientific and Cultural Organisation (Unesco). Thus the *South African Studies* CD-ROM entries are consistent with computerised cataloguing methods used internationally, in particular MARC (Machine Readable Cataloguing) that facilitates the exchange of standardised bibliographic data between countries.

The reason for having a standard bibliographic description is to ensure that there can be no doubt about which source is being referred to. Briefly, a bibliographic description would include some or all of the following elements:

> author
> title
> edition
> place of publication
> publisher
> date of publication
> pagination
> other details such as physical description, form, series, notes, standard numbers (the ISBN or ISSN)
> (Behrens 1994: 58)

When an individual library in South Africa adds a title to its catalogue, it first searches the South African co-operative library database (SADC) to determine whether the entry is

already listed. If it is, it will use the existing standard description in its own catalogue while at the same time adding its "holding statement" (a unique code identifying the individual library) to the entry in the SADC and so contributing to the joint national catalogue. Should the library find that the entry does *not* appear in the SADC, it would write its own bibliographic description in MARC[4] format (or, the South African equivalent - SAMARC, soon to be replaced by the USMARC to ensure an internationally recognisable exchange standard) and submit it to the SADC for inclusion. While the *South African Studies* CD-ROM occupies a traditional place within the established bibliographic record, the SALIT Web shares many of the features of a conventional bibliographic description. Indeed, the SALIT Web owes much of its core content to the South African National Bibliography (SANB), which was used (amongst many other sources) by Johan van Wyk in the compilation of our own database. Because the general features of standard bibliographic entries are included in its design, the SALIT Web can be used as a platform to access more detailed bibliographic information in the SACD by entering any of its bibliographic descriptors (author, title, subject) into an SACD search via SABINET (the South African Bibliographic and Information Network) and so tracking down a particular publication to library holdings in South Africa or abroad. Although the potential exists to access the SACD directly from the SALIT Web, there are several obstacles to achieving this ideal at the present stage of the project. One of the obstacles is the necessity of being a subscriber, via SABINET, to the SACD, and the two-step process of finding the reference in the database, and then logging on to SABINET to track down the library holding. At present this obstacle also applies to the user of the *South African Studies* CD-ROM, although the NISC is investigating the possibility of linking references to the user's own library holdings. By creating a new table in the database section of the SALIT Web, the existing bibliographic entries could be assigned MARC field names (e.g. Book Title could be tagged as the 245 MARC field). Following this route would greatly enhance the prospect of sharing the SALIT Web resources by becoming a participating member of the SACD. Once the full-

---

[4]   Machine-readable Cataloguing - a format created by the US Library of Congress so that library catalogue records could be read by a computer.

texts in the Library section of the SALIT Web are encoded in SGML according to the TEI guidelines (see Chapter 7), the same information would be required for each text "header" - a set of descriptions and declarations equivalent to the title page attached to a printed book. The TEI header elements have been designed to match MARC coding: the title tag in the TEI header is <titleStmt>, the equivalent of the 245 MARC field. (Giordano 1994: 391). The TEI Guidelines explain how TEI headers may be mapped (transferred by a computer) into MARC records (Sperberg-McQueen and Burnard, 1994: 672-76).

Its extensive bibliographical features, and its range across literature in all the South African languages has already made the SALIT Web the most comprehensive specialised bibliography in this field. The full-text and multimedia features of the SALIT Web set it apart from conventional databases, but its design incorporates the possibility of becoming a full bibliographical service in its own right at some future date. The following example may help to clarify the relationship between the SALIT Web and existing bibliographic databases:

A user of the SALIT Web, having found a reference to a publication would be able to scan the entry to discover how much information is available within the Database itself. If the search was for the author Thomas Pringle, the following screen would appear:

Figure 1: SA Literary Database Entry (early version): Thomas Pringle

Immediately available to the user are a short biography (upper right) which can be scrolled through, a lithograph of the author, dates of birth and death, and a sound recording of the first stanza of "Afar in the Desert" read by an actor. In the "Title" window, all the entries with Thomas Pringle as author are listed (in chronological order in a subsequent re-design of this search screen). The full-text entry for *African Sketches* indicates to the user that clicking on this item will load the complete 1834 edition of this work from the Library section of the SALIT Web onto the computer display for reading or for searching. However, should the user wish to locate a copy of the third entry, *Some account of the present state of the English settlers in Albany, South Africa*, (not as yet included in the Library section), s/he could use the author, title and publication details listed in the database

to query SABINET for the physical location of any print-based copies of the publication. By cutting and pasting the title into the SABINET search screen, the user would find the following entry:

```
        List of titles           Hits.... :  3
        Public Access Full Screen         Selected :  0
    _____

    |
    |
  1 | Pringle, Thomas, 1789-1834
    |   Some account of the present state of the English settlers in
    |   Albany, South Africa / by Thomas Pringle. - London : Printed for T
    |   & G Underwood, 1824. - iv, 125, [1] p. TITLE ID IS 65323
    |
    |   0730 X    DT 2400.A351 PRI!AFRICA
    |   4133 X    A 960.75 PRI
    |   6520 X    968.206 PRI!BAP
    |   9996 @@     @
    |
    _____

  Enter desired number(s), range, option  :
  Options : (?) for Help, (A)ll holdings, (#L)inks, (S)ort, (Q)uit, (CH)ange disp
        (#R)equest, (#H)olding.
        (F)orward
```

Figure 2: SABINET SADC database: Entry for Pringle's *Some account of the present state of the English settlers in Albany, South Africa*

The library holding codes indicate that the Jagger Library at the University of Cape Town (6520) and the Wartenweiler Library at the University of the Witwatersrand (0730) both have copies of the publication, and these may be accessed by the user through the Interlibrary loan system. This example shows how existing bibliographic databases may be used to complement the SALIT Web while highlighting the unique features of the Database itself. The example also points to the richer learning environment provided by the SALIT Web by giving the user a context in which to search for further materials.

Unlike the conventional database, the SALIT Web contains a wide range of possible directions for the user or learner, in which access to full bibliographic sources is only one such direction. Development priorities for the SALIT Web lie in supplementing the "rich mix" of titles, full-texts, reviews and multimedia items rather than in duplicating the efforts of the South African bibliographic community. By incorporating MARC records, the SALIT Web may become part of the mosaic of available bibliographic databases, but its true value lies in its integration of non-bibliographic resources.

An examination of the *South African Studies* CD-ROM user interface indicates some useful features which may be incorporated into that of the *SALIT Web*. Fourie and Behrens comment on the criteria used by reviewers of CD-ROM databases:

> Some reviewers focus on the database content and structure, while others emphasize the retrieval software, user interface or technical aspects. (Fourie and Behrens 1997: 116)

They go on to evaluate the CD-ROM under the headings of Vendor and Product details, Technical aspects, User interface, Searching and search management, Output, and Database content. The User interface and Searching and search management aspects of their review have relevance to a discussion of the *South African Studies* CD-ROM and the SALIT Web. They offer the following definition from Trenner of "user friendliness" as a basis for their evaluation. Trenner writes that a user-friendly system permits the user "to perform desired tasks without frustration and provides a range of features and functions which help each individual user to optimize his, or her, efficiency"(Trenner 1987: 105).

On this score, the *South African Studies* CD-ROM successfully presents a first screen that assists the first-time user by presenting a list of the eleven databases included on the disk together with simple directions for reading more about each one of them. Also included is the information that the search mode default is "Novice search" and how to switch modes to the "Advanced" and "Expert" modes. Once in the search screen, the user may invoke on-line help by pressing the F1 key, but there are also clear instructions at the bottom of the

screen that guide the user on how to proceed. The novice search mode also provides guidelines on using Boolean operators (AND, OR, NOT) and truncation (* for multicharacters and ? for single characters). Once a search has been carried out, the number of records retrieved for individual terms, as well as the total number of records retrieved are both displayed, and should the search produce a null result, the user is prompted to use the F1 key for guidance on search skills. The selection of search terms is aided by the addition of an automatic index that displays the closest match to the search term typed in by the user. The existence of duplicate records across the eleven databases is handled by a process which creates composite records, plus any additional data specific to a duplicate. "Output" refers to the way a user may handle the results of a search. Once a search has been completed, the retrieved records may be viewed in either short or long formats, the short format including only the publication title (and the journal title where applicable) and the long format providing all the details contained in the record. These results may simply be viewed on screen, saved (downloaded to a disk) or printed out.

The MS Access database used for the SALIT Web has all of the above features either as standard search tools, or as customised queries. The emphasis in the SALIT Web is to obtain a higher degree of user friendliness than is exhibited in the *South African Studies* CD-ROM, especially in view of the wide range of users at which the present project is aimed. A "Search Help" window, similar to that available in the *South African Studies* CD-ROM can be accessed from any of the views of the database, providing users with assistance with the simple "find" feature and providing guidance on using the various built-in search queries (e.g. all Zulu authors, or Women authors, etc.). From the primary "Author" search screen for example, the Find feature (represented by a binoculars icon on the main toolbar) can be used within the specific search field (Author) producing narrow results based on that specific field. Figure 1 (above) shows the result of an Author field search for Pringle. Should the user wish to scan the entire database for any mention of the author (or other element) across all the fields (Book or Article Title, or Commentary) s/he could use the Open Search form and by clicking the "Any field" choice, all records containing the search

word will be displayed. Either type of search would be appropriate for the novice user who can easily switch between forms, and so broaden the scope of the search. By following the latter search strategy (using the Open Search form) a user typing in the search term "Pringle" may find references in fields other than the Author field. The result of an Open Search, using "Any field" is presented in Figure 3 below:

Figure 3: SALIT Web Open Search form

In the above example, the user, having found "Pringle" in a book title, may click the "Filter by selection" button and come up with all the book titles that contain Pringle's name. By returning to the form and clicking the "Find next" button, the user may then discover Pringle's name in the title of an article, and once again click the "Filter by selection" button, producing a list of all articles containing his name.

As the SALIT Web developed, alternate access routes to the bibliographic information such as the Keyword feature were designed to provide guidance especially to inexperienced users. Using the current Beta 01 version of the CD-ROM, we will follow a limited number of possible search procedures, and in so doing illustrate the current organisation of the interface.

In Figure 4, the opening menu choices are represented schematically, with the
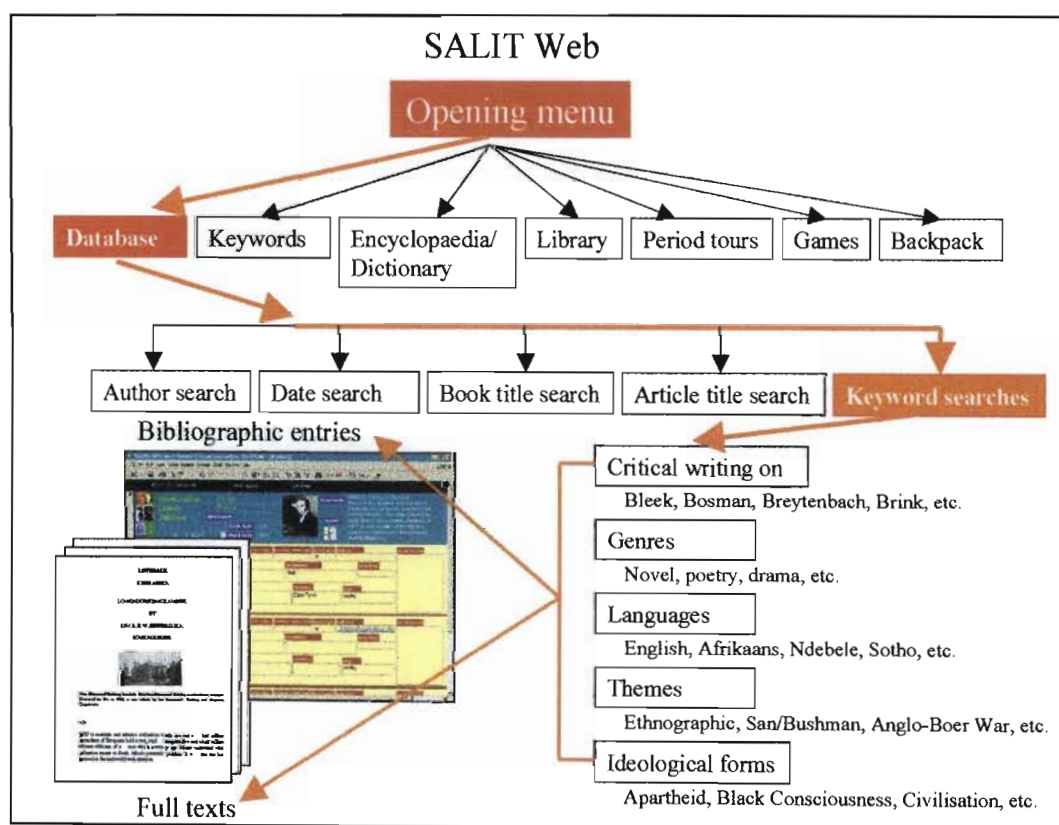
Figure 4: Overall structure of the SALIT Web

highlighted arrows demonstrating one likely route through the Keyword section to the full texts. Immediately a major challenge confronting the web designer becomes apparent in the appearance and layout of this page. Several different opening pages were experimented with
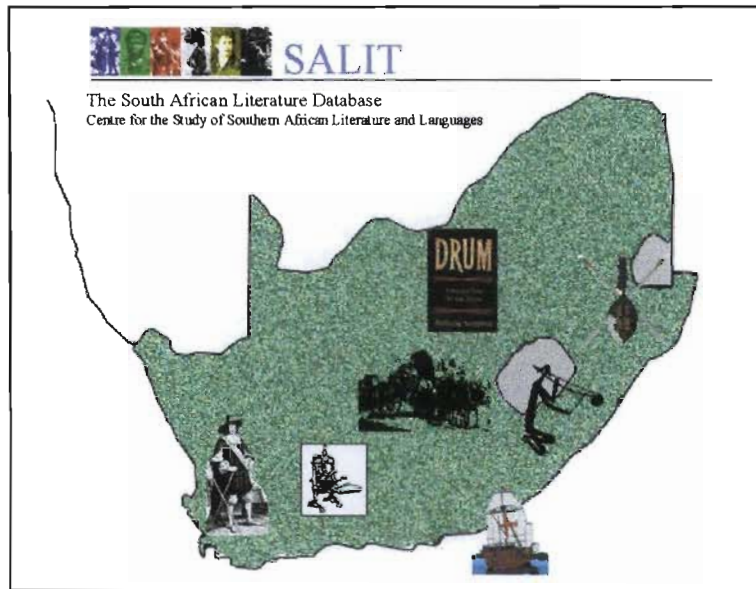
Figure 5: Alternate design for the opening menu
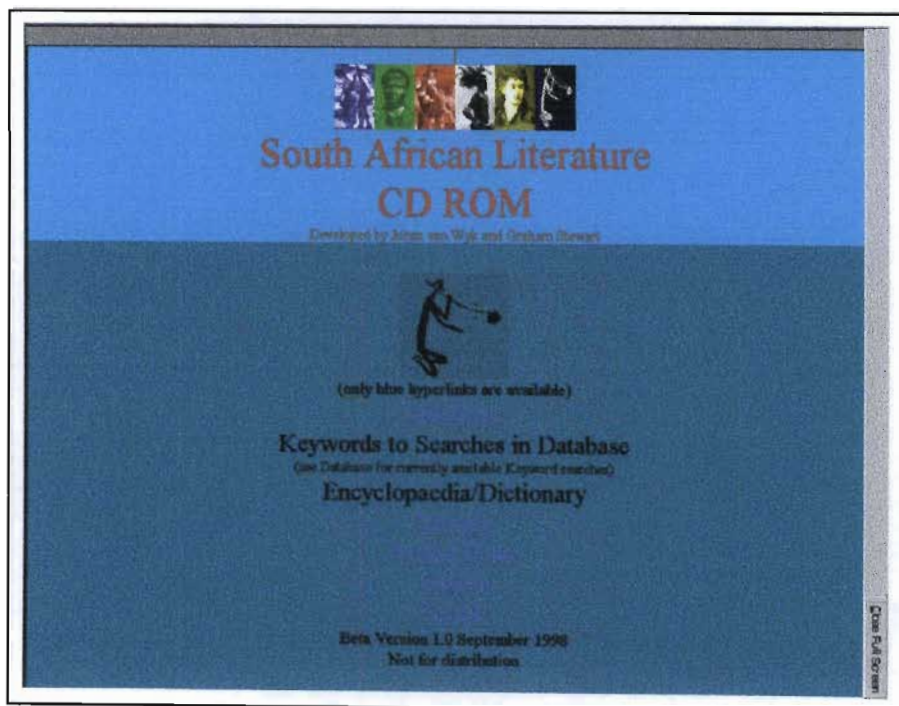
Figure 6: The opening menu of the SALIT Web

during the development of the web, including the map below (Figure 5), where a blend of geographical, chronological and historical icons present a very different conceptual impression of the web to the user than the one on the present CD-ROM. Clicking the image of the printing press would open a menu of references to the Lovedale Press and other links to South African printing and publishing, for example. The map metaphor has been used very successfully in the *The Atlas of Literature* (Ed. Bradbury 1996) and may yet appear as a navigation device as the web develops, but was considered too enigmatic for the students identified as the primary target group for this version of the SALIT Web. The non-sequential nature of hypertext would just as easily lend itself to an opening page that depicted page one of the full text of a canonical South African novel, encouraging the user to read on, relegating the options of searching the database, etc. to a less prominent position on a sidebar. Similarly, even the overview diagram of the SALIT Web (Figure 4, Figure 49) could be hyperlinked to the appropriate parts of the web, providing an opening menu that conveyed a much more compelling sense of the web-like structure of the resource.

The opening page finally chosen for the present CD-ROM version of the web (**Figure 6**) uses a textual menu more likely to be familiar to new users of the resource, and simply lists
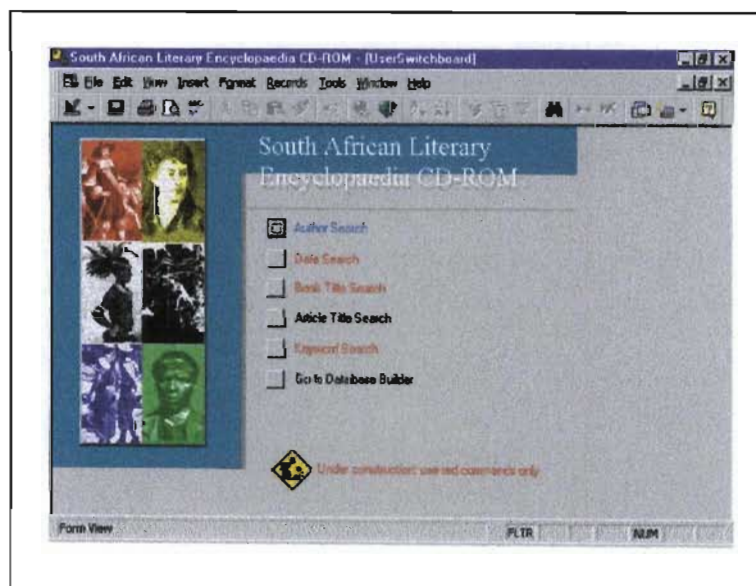
Figure 7: First page of the Database

various user interfaces: the database, a full list of key words (search terms), the Encyclopaedia/ Dictionary, Library, Period Tours, Games and Backpack. To illustrate its use, let us search for bibliographic information on Thomas Pringle by choosing the "database" option. The composite "open" search form illustrated in **Figure 3** is not directly available in the current Beta 01 CD-ROM version of the SALIT Web, now that the interface has been simplified to automate some of the more likely search strategies of a student user, hence the appearance here of an "intermediate" level of search options: Author search; Date search; Book title search; Article title search; Keyword searches (**Figure 7**).

Whereas the original "open" search form required that the user learn several steps such as identifying the Author field, selecting it, and then using the MS Access search button to perform the query, the new level leads directly to the five pre-arranged menus listed above. As already mentioned, the advantage of this design strategy is ease of use for the novice who is confronted with, for example, an alphabetical listing of author names, rather than a bewildering array of fields in which the position of the Author field may not immediately be
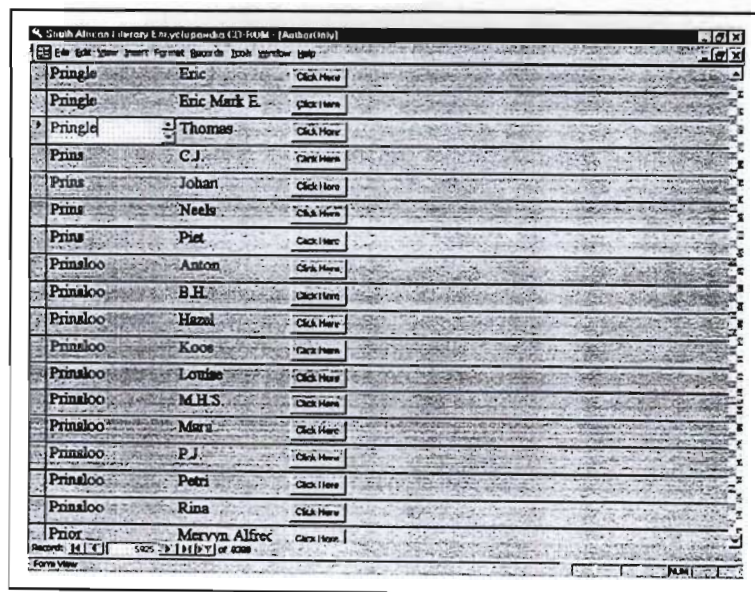
Figure 8: Author search screen

apparent. Of course, the disadvantage is the reduction in the number of formal field categories from which a more experienced user might make his/her selection at the start of

the search process. Formal items that do not appear here include publication details, ISBN numbers, and gender, but most of these fields become available once the user has reached the main bibliographic entry. The "open" search form will be re-introduced into later versions of the SALIT Web. Clicking on the Author option opens up the display shown in Figure 8.

Once the user has found the author's name, clicking the "click here" button opens the bibliographic form which contains the full list of the author's works, conventional bibliographic details, a photograph of the author where available, and an encyclopaedia-style biographical entry. Within the title listings are hypertext links to full texts in the Library section. Navigation between the bibliographic form and the full text requires only that the user click the Full Text hypertext "anchor" that in turn opens a new window containing the full text. By closing the full text window, the bibliographic display is restored. Returning to the main menu of the database can be accomplished either by selecting the "back" arrow icon in the body of the display or by closing the active window in the normal way. Bibliographic displays for any author together with all the functionality described above can be reached via any of the routes (date, book title, etc.) listed on the database main menu.

By far the most versatile of the choices offered on the database menu is the Keyword option. As I indicate in Chapter 3, the decision to index each bibliographic record with several different key literary concepts has elevated the query capabilities of the database to that of an encyclopaedia. Keyword coding of each record effectively brings to the SALIT Web a meta-classification based on the combined erudition and scholarship of the research team members. The resulting displays therefore link often highly disparate authors, titles and dates in terms of identified themes and ideologies, as well as the more conventional categories like genre and language. Janus (1997: 607) calls this a "blend of tradition and new technology" in which the accumulated knowledge of the literary scholar *cum* indexer propels the formidable processing capacity of the computer. All Literature webs share this general characteristic, but it is the topic or keyword classification that

exploits it most effectively. To take the idea further, the descriptive markup or encoding of full texts represents a similar combination of academic expertise and the computer's ability to sort, list and analyse with great speed and efficiency. Within the SALIT Web, there is categorisation at the macro level corresponding with the Keywords, and at the micro or document structure level, descriptive encoding of the full texts in the Library. TACT provides the means by which a user may "classify" or tag other structural features of the texts at a lower, more granular level. On the selection of the Keyword option in the database opening menu, the user of SALIT Web is presented with a selection of topics, categorised in the current version under the headings of: Critical writing on ... (individual authors); Genres; Languages; Themes; and Ideological forms.

Some of the keyword topics are shown in Figure 9:
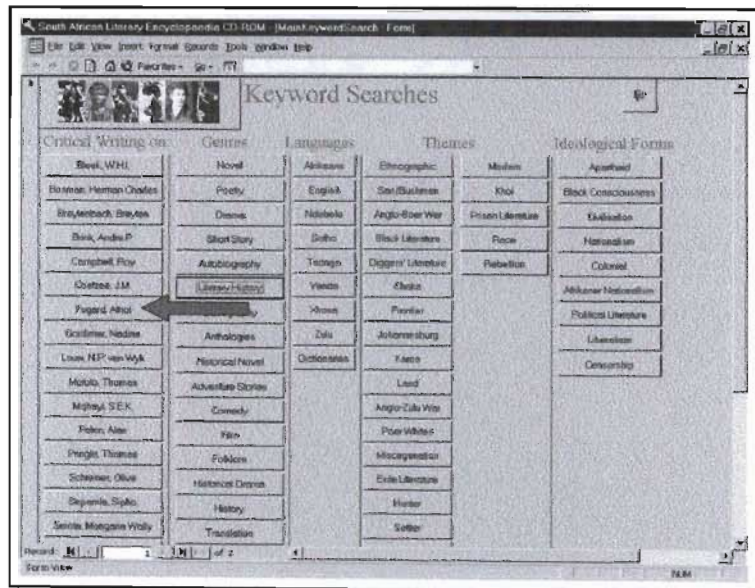
Figure 9: Keyword search form

Inextricably coupled with the expert direction available to the user through the keyword features, is hypertext's inherently democratic capacity to *resist* the primacy of any one voice in the text. The user is free at any time to ignore the pathways indicated by the keywords,

and should s/he discover any new categories, mark and link these her/himself and so construct alternate conceptual associations in the network.

One more of the Opening Menu choices is described here: the full text Library. Another, Period Tours, is covered in the next chapter, while other options on the menu are still in early stages of development and are not operational in the current version. The Library option offers the user access to the full texts without using the search features of the database and consists of an alphabetical listing of the titles and publication details. The benefit of the Library contents display is that it shows at a glance the current extent of the full text collection in SALIT Web numbering some forty-five titles at the time of writing. The title in the list is linked to the full text, and clicking the title has the same effect as choosing the Full text links in the database bibliographic displays described above - the full text is opened in a new window. Even as the collection list grows, it remains a familiar and accessible way of perusing the available titles. Most Literature webs and virtual libraries have similar displays which the user can scroll through, or perform simple searches with the "find" tool of a browser or other display application.

Among the texts currently accessible in the library are:

> D.F. Bleek, *The Mantis and his friends*
>
> W.H.I. Bleek, & L.C. Lloyd introduction to *Specimens of Bushmen folklore*
>
> Roy Campbell's article "Fetish worship in South Africa: A skirmish on the borders of popular opinion"
>
> Olfert Dapper's *Early Cape Hottentots*
>
> H.I.E. Dhlomo's "Literature and variety of tribal drama"
>
> James J.R. Jolobe's *Poems of an African*
>
> L.C. Lloyd *A short account of further Bushmen material collected*
>
> I. Nhlapo's pamphlet *Bantu Babel* on the language question from Edward Roux's 1944 sixpenny library series.

Olive Schreiner's *Trooper Peter Halket of Mashonaland*

By linking the bibliographic entry to the full-text, we have already exceeded the boundaries of the conventional bibliographic database. In the next chapter we examine the way in which the electronic medium has enabled encyclopaedias to challenge their traditional book-bound role.

Chapter 3 Encyclopaedias

The information retrieval industry has traditionally regarded encyclopaedias as useful, but essentially non-specialist sources. This applied particularly to "general" encyclopaedias like the *Britannica* but even to "subject" publications such as literary encyclopaedias that the user could scan for an overview or summary of an area before turning to more in-depth coverage in a specialist book or article. Some "dictionaries" and "companions" like the *Companion to South African Literature* are, according to this definition, also types of encyclopaedia. In the latter volume, entries are arranged in alphabetical order and include author entries with biographical and reference information, title entries giving a synopsis of canonical works together with a summary of critical opinion, and more general thematic entries on the Novel, the *Drum* movement and *Soweto Poetry*. Every entry includes a select bibliography pointing to significant primary and secondary sources for more thorough coverage of the material. Graphical material is limited to photographs of some authors, restricted to one or two per page rather than to every entry. The South African Literary Encyclopaedia in its printed and bound form would broadly fit the same pattern, but covering the full range of literature in all the South African languages.

This chapter will provide a short background to the development of encyclopaedias, describe common structural features and discuss the implications of the introduction of CD-ROM encyclopaedias since the 1980s. This leads to a discussion of the more recent phenomenon in a networked environment of the convergence of the four once distinct classes of publication: the encyclopaedia, the full-text database, the bibliographic database and the digital library.

The first encyclopaedia was the *Historia naturalis* produced by Pliny the Elder in 77 AD,

and included information on ethnology, geography, physics, physiology, zoology, botany, medicine, minerals and art (Behrens 1994:186-203). If "education" can be taken to include both "learning" and "research", the derivation of the term "encyclopaedia" from the Greek *enkyklios* (general) and *paideia* (education) corresponds with key elements of the SALIT Web, evidence of the convergence of the various information sources discussed below. Encyclopaedias were intended originally to encompass all human knowledge, and although that is an unattainable goal in a print-bound volume, it is interesting to compare the aspirations of the developers of the internet with those of first encyclopaedias. There is Ted Nelson's enthusiastic Project Xanadu vision (1987) of a site that would constitute a single "docuverse" where everyone had access to all published information at the same time, and the more down-beat, but significant declaration of Tim Berners-Lee, architect of the World Wide Web:

> HyperText is a way to link and access information of various kinds as a web of nodes in which the user can browse at will. Potentially, HyperText provides a single user-interface to many large classes of stored information such as reports, notes, data-bases, computer documentation and on-line systems help. We propose the implementation of a simple scheme to incorporate several different servers of machine-stored information already available at CERN, including an analysis of the requirements for information access needs by experiments (Berners-Lee & Cailliau 1990).

An on-line encyclopaedia today is potentially a universal information source. The World Wide Web (or Internet) already provides the platform but a different, more tangled web of copyright restrictions prevents its realisation.

Conventional encyclopaedias, then, provide a systematic overview of knowledge of either a general nature or in a particular subject area. In order to remain relevant sources of information, encyclopaedias must be regularly revised and updated, and the reputation of

major encyclopaedias depend on the frequency and reliability of revisions. We would use the *Historia naturalis*, still in print today, as an historical artefact, not a source of current information. Reputable general encyclopaedias like *Britannica* and the *World Book Encyclopaedia*, are subject to continuous revision but it is illuminating to examine the revision policies of publications in more detail. The *Britannica* is printed three times a year, but these revisions are to certain selected entries only and the last completely new edition was the 15th edition that was published in 1974. Also of considerable importance in the overall design of an encyclopaedia is the intended target audience. This will influence the selection of entries (in say, a general as opposed to a specific subject, encyclopaedia) and also the accessibility of the language used in the entries depending on the age group of the intended reader. The *Britannica* is aimed at an adult audience, whereas the *World Book Encyclopaedia* is designed for senior school pupils, but is written in a straightforward style that would also be suitable for non-mother tongue speakers of English (their advertising copy stresses that each entry is "written in the familiar *World Book* style - clear, direct, easy to understand, and in International English"). Accessibility to a wide South African audience, the majority of whom do not have English as a mother tongue, is another key feature of the SALIT Web implying that at the level of the literary biographies at least, we use a plain English policy that avoids jargon and includes clear explanatory material. The SALIT Web full-texts are not confined to English, though. As the collection grows, it will include titles in all the South African languages from the CSSALL's translation projects and African Language departments at other universities (see Chapter 9 for more on this development). The hypertext environment can easily provide links between the text in its original language, and a translation as can be seen in the illustration from the Period Tour, below.

Encyclopaedias are usually arranged in an A-Z alphabetical manner and include a comprehensive index. The SALIT Web conforms to these criteria through the provision of sorting features that list either author names or titles in alphabetical order. In addition, of

course, the Database like most on-line information sources may also be viewed by classification, e.g. chronological order, or by genre or language. Print-based encyclopaedias attempt a similar form of accessibility by providing time-lines and summary articles, but neither of these matches the searching power in an electronic medium. The *Britannica* has 32 volumes made up of a two-volume *Index*, a one volume *Propaedia* (divided into subject areas with overviews and references to in-depth information in the other volume), a twelve-volume *Micropaedia* (containing short, paragraph length entries arranged in alphabetical order) and the seventeen-volume *Macropaedia* that provides the most thorough level of information, also arranged in alphabetical order. Within the longer *Macropaedia* articles, reader "navigation" is facilitated by the use of a mini book structure that breaks the articles under sub-headings that are listed like a "contents" page at the start of the entry. Further references and key words appear in the margins alongside the main text. The progression in the depth of information from index to general field, to summary entry to in-depth article (with a chronological/time-line option providing an alternative dimension) is a reliable and time-tested knowledge structure that should inform the final interface to the SALIT Web.

The electronic *Britannica Online*, based on the print version follows roughly the same type of arrangement, although the volume titles have disappeared, replaced by buttons that point the user towards "index", "articles" and "knowledge in depth", amongst others. The views offered by these buttons approximate the *Index*, *Micropaedia* and *Macropaedia* respectively, while a "spectrum" button leads to the on-line equivalent of the *Propaedia*, described here as "*Britannica's* outline of knowledge with links to relevant articles". Further buttons provide access to sub-sets of the above information under the headings "biographies", "geography", "illustrated articles" and "multimedia".
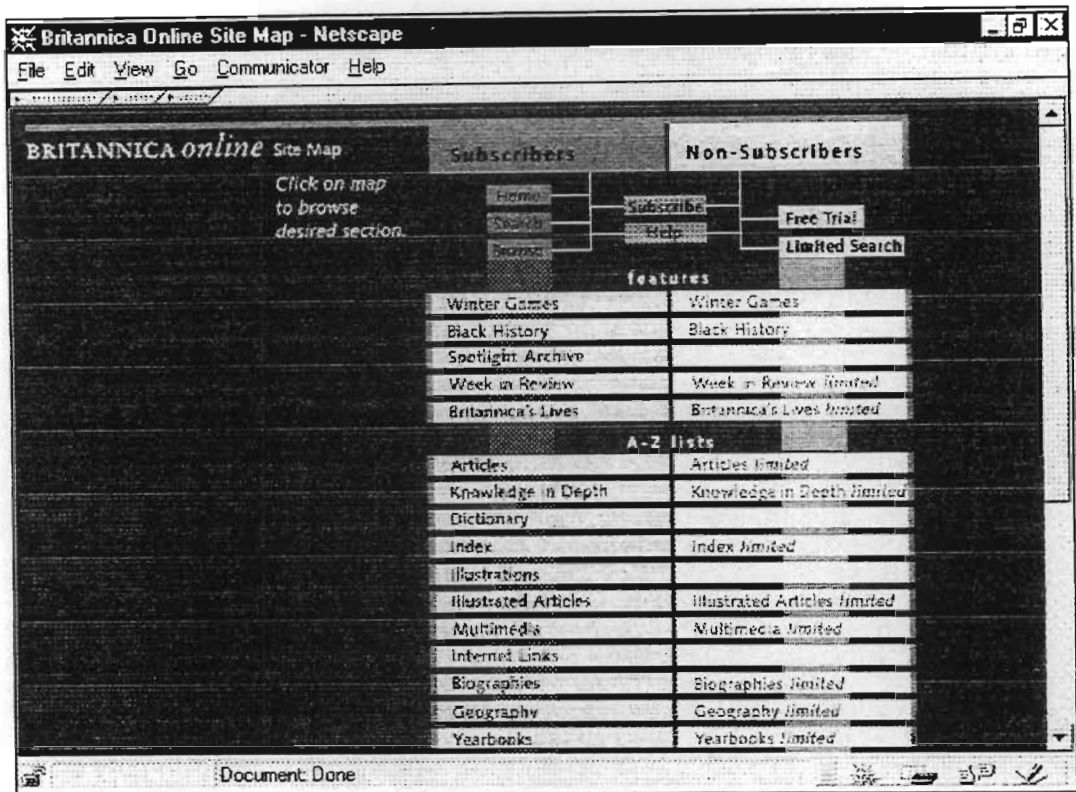
Figure 10: *Briannica Online* Internet gateway

The history of CD-ROM encyclopaedias began in 1986 with the appearance of the Grolier *Electronic Encyclopedia*. It was a development on CD-ROM of what was originally a full-text database of the *Academic American Encyclopedia* that had at first only been available to subscribers via on-line networks. With the recent advent on-line of a giant like the *Britannica*, this is an interesting reversal - and it seems inevitable that the Grolier will eventually also become available on-line again, but this time with colour graphics and multimedia unlike its text-based predecessor, and on the World Wide Web. (The Grolier has already re-appeared in a computer-based multimedia format as the *New Grolier Multimedia Encyclopedia*.)

While the CD-ROM encyclopaedia's greatest strength lies in its powerful search, retrieval

and linking capabilities, its main marketing feature has been the introduction of multimedia. Multimedia integrates text, images (either photographs, diagrams or maps), animation, digital video and sound, making it possible for the user to move, as in the *Britannica*, from a feature article on Shakespeare to contemporary illustrations, recordings of familiar speeches, 3-D models of the Globe Theatre and video clips from the plays. The Microsoft software company has quickly stepped into what used to be a publisher's domain by purchasing the 29-volume *Funk and Wagnalls New Encyclopedia* transforming it into the *Encarta Multimedia Encyclopedia* and marketing it as a free accessory with their Windows operating system.

By far the most interesting development of an encyclopaedia in CD-ROM from the point of view of the present SALIT Web project is *The World Book Multimedia Encyclopedia*. Although it is not yet available on-line, the Online Computer Library Centre (OCLC) bibliographic database service has announced its imminent inclusion in its "FirstSearch" Electronic Collections Online service (Smith 1997). Originally, *The World Book Encyclopedia* was designed to meet the needs of the American primary and secondary school curricula and therefore tried to make the entries as accessible to this age group as possible, both in structure and in language style. Since the 1992 International Edition, the content has widened sufficiently for it to be classed as a full "general" encyclopaedia similar to the *Britannica* in scope, if not in depth. The print version's twenty-one volumes are arranged in alphabetical order, incorporating its levels of information within each entry, progressing from straightforward definitions to more challenging articles, unlike the *Britannica*, which divides its levels of information into completely separate volumes (see the description above). Tracing the development of the CD-ROM version in "From Paper and Ink to CD-ROM: Digitizing the *World Book* Image", Robert Janus (1997: 604) comments that "the book metaphor does not always succeed on a computer screen" and shows that the CD-ROM version is a re-interpretation of the original encyclopaedia in the light of new technologies, rather than a faithful reproduction of the printed volumes in digital form. The

first CD-ROM was published in 1990 as the *Information Finder*, a text-only version based on the database that was used to produce the print version. There is a close parallel here with the development of the first on-line bibliographic databases that also started as computer tools for the production of printed books, but came into their own as on-line publications in their own right, and are replacing the books they once helped produce (see Chapter 2). In several ways the transition from print database to on-line *Information Finder* matched that of my own development of the SALIT Web from Johan van Wyk's original *Concise Historical Survey: South African Literature* (1996b) which was intended to give a diachronic account of South African literature (Chapter 10 contains a detailed account of the conversion of the *Survey* document to a database). Where the *Survey* had to be stripped of MS Word "procedural" markup codes (font, italics, etc.) the *World Book* database contained over 400 codes, more than 300 of which were "procedural" markup that had significance only for the page composition process. Only about ninety codes in the *World Book* database were useful for electronic display and retrieval, so the unnecessary (and undesirable - they had the capacity to crash the retreival database) codes had to be laboriously removed. As with the *Concise Historical Survey: South African Literature*, not all of the process could be automated forcing the *World Book* developers to find manually, for example, all instances of sentences that did not end with a full-stop and two spaces, as these would be incorrectly interpreted by the retrieval database as field endings.
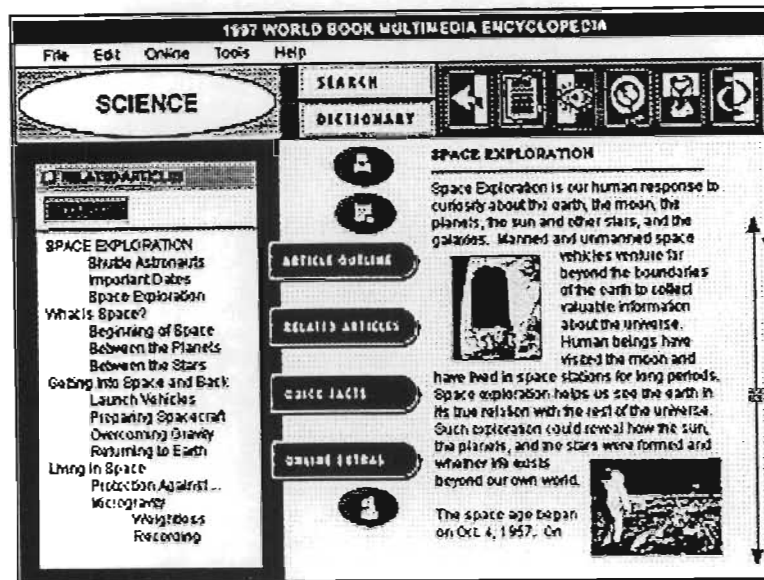
Figure 11: *World Book* - sample entry page

A unique (and proprietary) search engine was developed by *World Book* to combine

conventional Boolean searching with topic searching. The inclusion of topic searching was a

priority because it took into account the likely search behaviour of the target users - primary

and high school children - who might be unable to select the information they were looking

for from the type of list that a purely Boolean search would provide. Typically, a Boolean

search would use the search term (say, "liberation struggle") to retrieve all instances of the

word string. Not all of these would necessarily be what the user was trying to find. In the

topic search engine, the user is taken straight to an entry on the Liberation Struggle and

with another mouse click display several related article titles on, for example, Worker

Poetry, or biographical entries on specific writers or book titles. The key to the *World Book*

topic search was the existence of an index, prepared by professional indexers, that already

contained directions to related material under each entry. By linking the index to the search

engine, the engine first consults the index, which then points directly to a pre-indexed set of

links, starting with the one that most closely matches the search term.

> The user can select any or all of these additional hits, assured of their pertinence because a
> human intelligence has already associated each of these topics with the search term during
> the indexing process. (Janus 1997: 607)

From the start of its development the SALIT Web has contained an indexing element that
can be used to develop a similar "topic search" facility. The "keyword" field, containing
nearly 100 separate index terms ranging from genre identifiers to references to discursive
formations is embedded in the structure of the database and provides a platform for a full
indexing project that will ensure greater accessibility to the entries, both for the novice and
the more experienced user. The pilot version of the *Database* has the capacity to provide
Boolean searches across all fields and within particular categories (e.g. author and title) but
even with the present set of keywords, certain topic searches could be constructed by
directing a search to the keyword field, and using that to list entries containing the term. To
fully exploit the topic search capability, however, would require an additional indexing
project. The topic search option in the *World Book* is actually different from another finding
aid, the "Infotree". Built on an existing breakdown of titles into subject areas (for assigning
updating responsibilities to subject editors), this facility allows the user to browse through
articles by content, from eight primary headings including "Humanities", "Social Science"
and "History". There is no direct equivalent of this breakdown in the SALIT Web because it
is a "subject encyclopaedia" rather than a "general" one.

Once the decision had been made to produce a multimedia version of the *Information
Finder*, many of the design questions that had to be answered were precisely those that
applied in the case of the SALIT Web:

> How many pictures? Which ones? How big would they be? How sharp should they be?

How many maps? Which ones? How detailed should they be? [...] And each answer
claimed its own bit of "real estate" on the CD-ROM. (Janus 1997: 616)

Copyright issues loomed large in this stage of the *World Book* project. Although many of

the pictures used in the print version of the encyclopaedia were owned by the company, new

licence fees had to be negotiated for electronic publication of those they did not own. And

in most instances, the right to print such pictures was not available. Rights to print almost

every picture in the SALIT Web CD-ROM archive will have to obtained before it can be

publicly distributed, representing a major task of documenting each picture, tracing the

copyright holder, and requesting permission to re-print it electronically. This may not be as

formidable a problem as it at first appears, as many of the images in the SALIT Web gallery

belong to State organisations which usually require only the granting of permission.

Copyright, and its implications in realising a project of this type cannot be underestimated.

The spectre of the Boazs' T.S. Eliot CD-ROM looms above current projects in this area:

although focussing on *The Wasteland*, and with all the multimedia possibilities open to

them, the authors could not obtain (cost-effective) permission to use either the full text of

the poem or of Eliot's own recorded voice reading it (Boaz & Boaz 1996: 133).

Another feature of the *World Book* that could most profitably be replicated in the SALIT

Web is the "Timeline" feature (see Figure 12). The many specific time-lines contained in the

original text version of the encyclopaedia were used as a basis for the CD-ROM version that

allows the user to recognise the time-relatedness of events and people. The timeline also

contains selected graphics from the graphic database, and clicking on references or pictures

in the time-line takes the user to a related article.
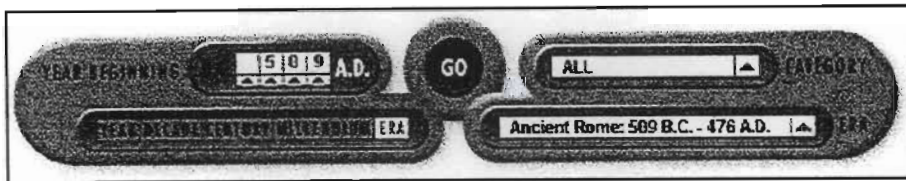
Figure 12: *World Book Encyclopaedia's* Timeline feature

At one level, it can be seen that the SALIT Web *is* an encyclopaedia, while we have discovered in the previous chapter that it also bears a close relationship to a bibliographic database. Networked information sources have blurred the once clear boundaries between the encyclopaedia, the full-text database, the bibliographic database and the digital library. Established "names" in each area (*Britannica* and the OCLC) now have on-line manifestations that share or completely duplicate the types of information resource once confined to the printed and bound versions of the other publications. *Britannica On-line* has an Internet gateway with the look and "feel" of the familiar multi-volume set, with the same content, but with the capacity to direct its users to full bibliographic databases and full-text versions of specific references. Once a select bibliography is hyperlinked to resources outside the boundaries of the encyclopaedia itself, it becomes, essentially a gateway to a virtual library. Similarly, the OCLC, a reputable on-line bibliographic service is extending its links beyond bibliographic catalogues to full-text on-line journals and periodicals providing several levels of information: from the bibliographic reference itself, to full-text abstracts and on to the complete publication. The knowledge structure underlying this arrangement is essentially the same as an encyclopaedia: alphabetical head entries, overview/summary information and then full-text books or articles. K. Wayne Smith, President of the OCLC predicted that all forty-eight of their Electronic Journals Online (EJO) titles would be available on the internet by the end of 1996, and that the company was continuing to develop its electronic, scholarly publishing operation (Smith 1997). He mentioned two other initiatives, Elsevier Electronic Subscriptions (EES) program in association with Iowa State University and the University of Toronto, and the "Red Sage" project, run by the University

of California at San Francisco, both of which designed to be "electronic libraries in scientific, technical, and medical disciplines to deliver journal information to end-users' desktops".

This is a virtual library by another name. An example of an on-line digital library undergoing a metamorphosis in the opposite direction - from a set of full-text resources to summarised overview entry points - is the Library of Congress. Here, a new gateway providing "pathways" classified under headings such as "People" and "Events" guides the user towards digitised full-text and multi-media resources and archive material (see Chapter 5 for a description of this resource). In the SALIT Web, the Period Tours option offers similar pathways for the user. One such tour, the "0000-1652 (From Time Immemorial to the arrival of Van Riebeeck)" is illustrated in the next chapter, which deals with Literature Webs, hypertext environments specifically designed for the exploration of literature.

Chapter 4 Literature Webs

In the previous chapters on Bibliographic Databases and Encyclopaedias, we have seen the electronic resources developing out of what were essentially traditional linear book-based resources rooted in the Gutenberg era, albeit with the potential for providing information in branching and interlinked ways. The natural successors to these hybrid texts are hypertext webs that have their origin within the architecture of the new multi-layered computer network environment - and are products both theoretically and practically of the Information Age. The SALIT project is a species of this new genus, and like the networked resources described below, has an intention and a form that owes more to the enabling technologies of information science and the notion of its relative place within the "general economy" of discourses than to the limited texts of the bibliography or encyclopaedia.

Brown University in the United States has been the site of significant developments in the use of hypermedia for scholarly purposes in the humanities, and houses two of the most important literary/historical sites in the English language: The Brown University Women Writers Project and The Victorian Web. As can be seen from the entry in the Hypertext Timeline (Chapter 1, page 19), Andries Van Dam, a member of the Brown University Intermedia Project (Welsch 1992: 626), was one of the first to use hypertext/hypermedia in literature. In an experimental course based on William Blake's "The Sick Rose", students were able to refer to related materials from the node containing the poem, as well as participate in the shaping of the materials. "The Sick Rose" project used many of the features common to present-day hypertext documents: multiple windowing, movement "outward" from the text of the poem itself, random access to related texts and the concept of the poem as part of a "web of interconnected materials ... [in which] each student could use the strands of the web to "pull" other selected materials into the three remaining

windows" (Catano 1979: 270). Interestingly, Catano also notes the extent to which the computer activity enhanced student creativity and classroom discussion rather than supplanting the lecturer as might have been expected.

The Victorian Web developed out of another application, *Context32*, a part of the Intermedia Project aimed at history of literature, or literature survey courses. In the case of "The Dickens Web", 2.5 megabytes of texts relating to Dickens himself and specifically to *Great Expectations* were linked together in more than a thousand information nodes. The project attempted to provide a context for the literary work, and reduce the didactic relationship between lecturer and student. The web provided a flexible learning experience for students by linking the work to its contextual surroundings in historical and biographical data. The following illustrations (Figure 13 and Figure 14) from the Dickens section of the later Victorian Web give an idea of the interface used.
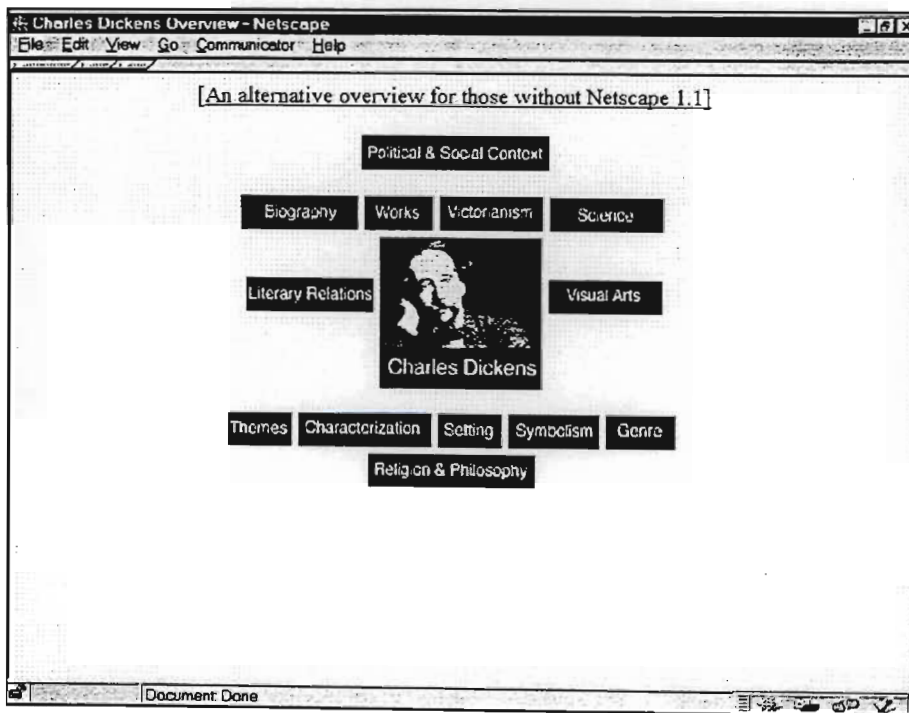
Figure 13: The Victorian Web - Dickens opening page

Clicking on, for example, "Biography", retrieves another node (page) containing links to biographical material ranging from a Chronology to socio-historical commentary relating to the publication of his books. Graphical anchors point to full texts of some of his novels available within the web: *Great Expectations* and *Little Dorrit* (Figure 14).
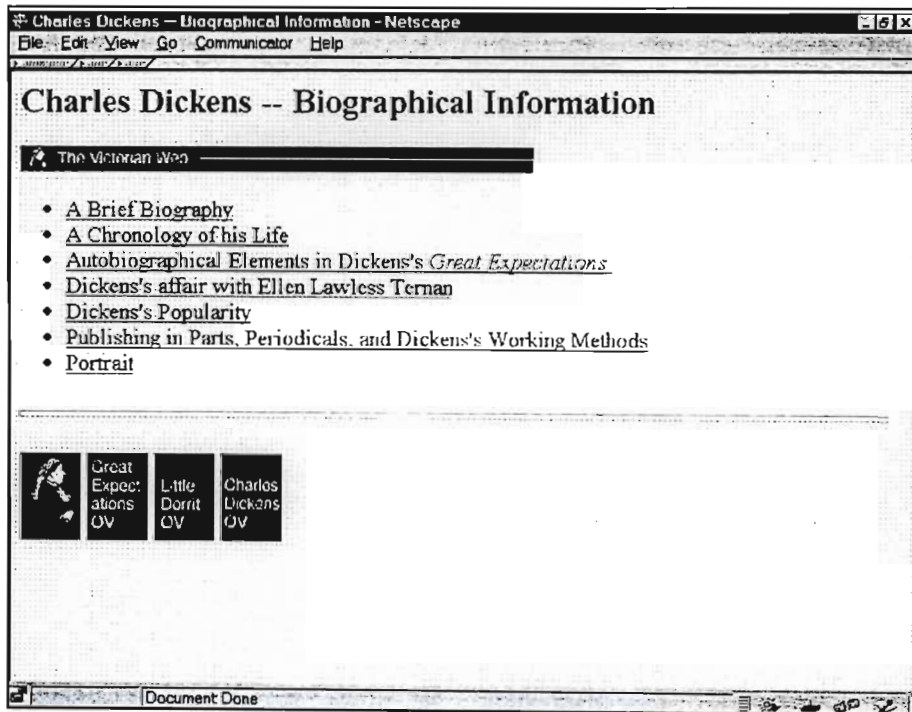
Figure 14: The Victorian Web - Dickens Biographical opening page

Landow, a Faculty Fellow at Brown University's Institute for Research in Information and Scholarship (IRIS) from 1985 to 1992, worked as a member of the team that developed Intermedia. He supervised, edited, and partially wrote various hypermedia documents on this system used to support English courses ranging from introductory surveys to graduate

seminars. He has published the *Dickens* and *In Memoriam* Webs in Storyspace (Eastgate Systems 1992) and *Writing at the Edge*, a collection of Brown student Storyspace webs. He created and maintains the Victorian Web (into which the Dickens Web has been absorbed) and is a much amplified WWW version of materials originally created by Landow and his colleagues in Intermedia and Storyspace. The "overview" page (Figure 15) contains links to "Victorianism" and "Literature", in which twenty-five further "webs" in the pattern of The Dickens Web lead to resources relating to each of the authors, including Lewis Carrol, Charlotte Brontë, George Eliot and Elizabeth Gaskell. (Readers of the hypertext version of this dissertation may use the Victorian Web overview page to link directly to the Brown site via their Internet browsers).
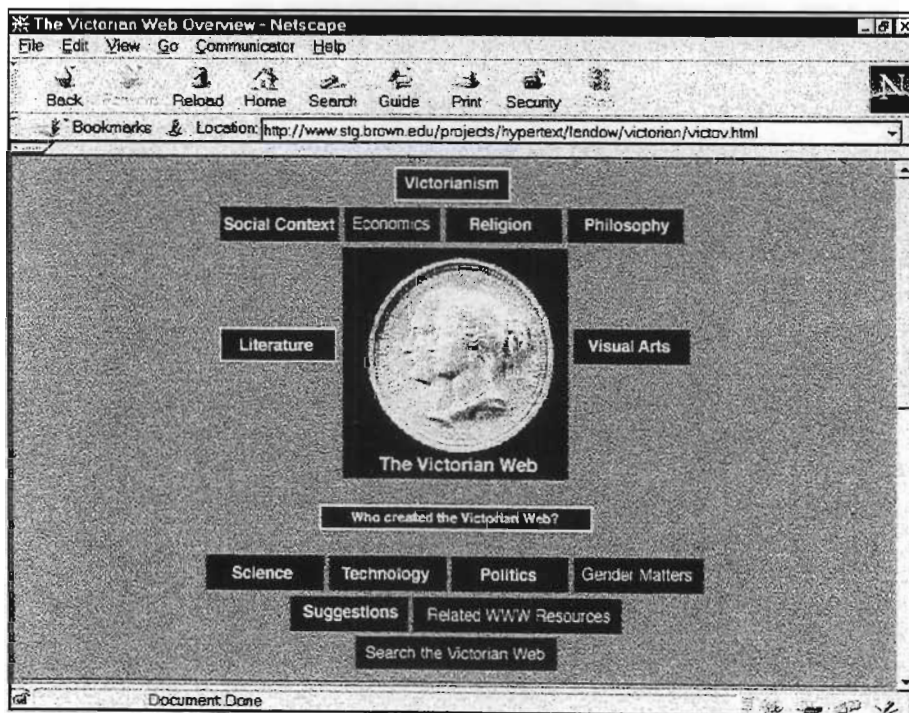
Figure 15: The Victorian Web - main opening page

Overview pages, like a table of contents, provide a convenient entry point for the user. There is no need for the overview to take the linear form of a contents page, however, nor does it have to point only to resources within a particular "book". Recognising that

hypermedia, while providing efficient ways of pursuing traditional scholarly activities, like searching for sources or scanning through journals represents a fundamentally new way of relating to information, Landow offers a set of rules appropriate to what he calls the "rhetoric of hypermedia" (Kahn 1991: 251). The nineteen rules imply an epistemology or knowledge structure unique to hypertext, for example the importance of links always expressing connectivity between anchors, whether the relationship is explicit (as in a direct reference to another text) or implicit, as in a glossary explanation. Other rules suggest conventions that express both the connectivity and the multiplicity of relationships to be found in the materials at hand.

Hypertext webs may be classified as either or "axial" or "networked" in structure. The axial structure places an individual work at the centre of the web, with additional information in the form of commentaries, annotations or variant readings radiating from the centre like the branches of a tree. It is this axial structure that has been adopted for the further development of Pringle's *African Sketches* and the CSSALL's new *Dube Web* project, based on translations into English of the writings of John Langalibalele Dube, early 20th Century politician and founder of the newspaper *Ilanga Lase Natal*.

Pringle's *African Sketches* (1834) in the SALIT Web restores the original format of the work: combining Part 1: *Poems Illustrative of South Africa* and Part 2: *Narrative of a Residence in South Africa* into a single electronic text. A full word index allows quick searching for words and phrases and allows cross-referencing of poems, notes and the text of the *Narrative*.

Thomas Pringle's 1834 text of *African Sketches* was selected as a suitable pilot project in which a significant South African text could be located at the centre of a hypertext web. In addition to its being one of the over 35,000 "title" references in the SALIT Web

bibliographic database, it was designed to show how each SALIT reference had the potential, over time, to be linked to the full text to which it referred. The design of the electronic publication followed as closely as possible the original print-based book - cover pages and front matter, Parts 1 and 2 of the text, the contents, major section divisions, footnotes and plates. The database provided a hypertext link to the title, and from there on, each of the elements listed above were hyperlinked within the micro environment of the electronic web structure.

The Pringle text offered considerable scope for the development of a wider information web around the original text, including other works by Pringle, contemporary documents of historical and social significance, audio and photographic information, maps, commentaries, reviews and criticism. Another promising avenue for the assemblage of an information web was Pringle's involvement with the establishment of South Africa's first newspaper, *The South African Commercial Advertiser* and the subsequent bitter conflict between the publishers and the governor of the Cape, Lord Charles Somerset. As I mention later in this discussion, newspapers have played an important part in the development of South African literature and politics, and as such could form a valuable thematic link across the SALIT Web as a whole.

The process of producing the e-text - starting with the origination of the text (scanning) through later stages of proofreading, editing and setting within a hypertext environment - proved more complex and time-consuming than anticipated. The complexity resided not only in the technical challenges posed by the digitising of the various documents and visual material, but also by the conceptual challenge of valid electronic representation of textual and meta-textual elements (like authorial footnotes and biographical material), developing search criteria, classifying and identifying hypertext links (for example, discriminating between simple explanatory pop-ups containing later critical commentary or the author's

own footnotes and endnotes). The relationships amongst these various meta-textual elements now present themselves as a clearly-defined system to the reader/user and although some of the more familiar relationships have an established place with reference to the original text - like the footnotes and endnotes - others do not. For instance, *African Sketches* has a chapter consisting entirely of biographical notes on some of the personalities referred to in the body of the text. The scope of these biographies is constrained by the 1834 publication date, and fuller details (including the date of death, for example) required clear demarcation as later editorial additions. The most appropriate way in which to incorporate these elements had to be considered within the conventions of scholarly publication while taking into account factors based on hypertext theory such as ease of navigation and coherence in an information web.

The present electronic version of the *African Sketches* text allows the reader the same access to the material as his or her 1834 peer: i.e. all the poems in the *Poems Illustrative of South Africa* section and the *Narrative of a Residence in South Africa*, together with all authorial notes, footnotes and endnote material. Beyond the more obvious differences in the medium that the present-day reader of the electronic version would use (the computer screen rather than the printed page) there are more profound differences. The electronic version of *African Sketches* can be read like any conventional book, but may also be used as an interactive search and discovery medium. Should the reader wish to enquire about the nature and quantity of Pringle's references to women, particularly black women, s/he could perform several searches - for example, for the occurrence of the word "woman" within five words of the word "black" throughout the 115 poems and 350 pages of the two texts - and this would take less than 30 seconds. The textual analysis of the Pringle texts is covered in greater detail in Chapter 6.

Because additional material has been added to the broader environment of the database,

detailed maps and photographs of the Eastern Cape and the Baviaans River Valley are also available to the interested reader. The journalistic possibilities around the *South African Commercial Advertiser* theme have not yet been developed, but facsimile copies of some pages of the paper have been included as well as contemporary illustrations and documents related to Pringle's own account of his conflict with Somerset in Chapter 10 of *Narrative of a Residence in South Africa*.

The Folio Views software used to view and navigate through the *African Sketches* text is designed to encourage interaction between reader and text, not only through hypertext linking and speedy search and retrieval, but through being able to add one's own annotations and material to the texts (either temporarily or as a permanent addition to the electronic text). This option opens up considerable scope for using the *African Sketches* text for further research and for learning. A scholar exploring Pringle's references to, say, "work" and "economy" would be able to add permanent links to the text, as well as incorporate her own notes, commentary and related articles to the original web of information. In a very real sense, the reader/learner becomes a co-editor of the text and the centre of another web of references customised for the use of another set of researchers or learners.

Both the *African Sketches* and the Dube webs retain a sense of a unified main text, with the central work providing an axis to which other information, textual and graphic, remains supplementary. Kolb identifies this structure as one the traditional uses hypertext could perform in philosophy:

> It could present a reference collection of philosophical texts that included links from texts to commentaries, discussions, alternatives, refutations, bibliographies and the like. (Kolb, 1994: 324)

However, as Landow points out, once there are a number of these axial structures, it makes sense to re-use some of the information that might relate to more than one of them - in the case of the Pringle and Dube texts, the role of newspaper publishing in South African politics would be an obvious link - and hence a process of networking the two axial webs would have begun. The networking of information is a key component of the SALIT Web, where the linking of bibliographic information to full texts already provides one route or pathway, while the keyword "newspaper" would lead the user to both authors. The networked structure is also apparent with the development of the "dictionary" section of the SALIT Web, where the dictionary, as part of the overall web structure, is available to the user of the Pringle, Dube or any other of the full texts in the virtual library.

Some of the shortcomings of networked webs are apparent in the Brown Postcolonial and Postimperial Web. Constraints such as copyright restrictions and the sheer enormity of task of collection building for a comprehensive virtual library can leave the serious gaps in what on the surface appears to be a rich resource. In the Brown Web, the links are largely to bibliographic details and general historical and literary background with very few full texts: the full text "Xanadu" ideal in which the web could be extended to current publications via a pay-as-you-use system has yet to be realised. An exploration of the "Union (sic) of South Africa" site within the Postcolonial and Postimperial Web reveals that below the surface of the overview page, there is information on only one South African author - Nadine Gordimer - and only one of her works - *July's People*. The pages are not complete, but contain some commentary and background on the novel. Other anchors on the "South Africa" page, for example "Bibliography" and "Literary relations" are not currently linked to any information at all. Other nodes contain information gathered from public domain information sites: "Communication" (from the CIA) and "Music" (from Christian music in South Africa). Some of the available African authors are listed below (blue underlined entries are anchor links, but the Jean Marquand (sic) link was not operative):

Muhammed ben Abdallah (Ghana)

Chinua Achebe (Nigeria)

Catherine Obianuju Acholonu (Nigeria)

Zaynab Alkali (Nigeria)

Bruce Chatwin (U.K.) [under construction]

Kamala Das (India)

Anita Desai (India)

Tsitsi Dangarembga (Zimbabwe)

Cyprian Ekwensi (Nigeria)

Buchi Emecheta (Nigeria)

Chitra Fernando (Sri Lanka)

Shiela Fugard (South Africa)

Nadine Gordimer (South Africa) [under construction]

Bessie Head (Botswana

Farida Karodia (South Africa)

Jean Marquand (South Africa)

Ezekiel Mphahlele (South Africa)

Charles Mungoshi (Zimbabwe)

Njabulo Ndebele (South Africa)

Ngugi Wa Thiongo (Kenya)

Isidore Okpewho (Nigeria)

Ben Okri

Michael Ondaatje (Boston University site)

Tess Onwueme

Femi Osofisan (Nigeria)

Sembene Ousmane (Senegal)

Salman Rushdie (India & Pakistan)

Ken Saro-Wiwa (Nigeria; Ogoni)

Wole Soyinka (Nigeria)

Zoe Wicomb (South Africa)

Musaemura B. Zimunya (Zimbabwe)

(Brown University Internet Site: Postcolonial and Postimperial Literature Web)

Much more successful than the embryonic Postcolonial web at Brown is the well-established Women Writers Project. The original objective was to produce a single scholarly anthology that asked new questions about genre, publication, canonical traditions and literary culture. The strong literary computing legacy at Brown turned the founders of the project to computer technology as a natural tool for providing efficient and innovative access to this material. As the project developed, its aims were expressed as: (1) produce scholarly anthologies and individual volumes of pre-Victorian Women Writers; (2) contribute to research on literary databases and text management systems; and (3) serve as a central research resource with a full-text electronic database of pre-Victorian women writers in the English language. As can be seen with the "Victorian Web", the literary orientation of the computer technology was already in place as a foundation for such a project.

The Women Writers Project is an important model for the development of literary "textbases" in South Africa for two reasons: firstly, by building an electronic collection of women's writing in English between 1330 and 1830, material which has been largely inaccessible to researchers and students is becoming available for study; and secondly, by making previously marginalised texts available, the project is able to challenge the distorted view of the role of women in Western literary and cultural history. The parallels with the suppressed and marginalised literature of our past are obvious, and the Women Writers Project offers a practical way in which to collect and make accessible important texts that have long been out of print, or perhaps never formally published. The University of the Western Cape's Mayibuye CD-ROM project has already had success in this respect. Where the Women Writers Project textbase disseminates its texts through on-demand publishing, traditional print publishing of selected works, provision of selected texts to database publishers, and Internet access - the Mayibuye project has bundled print copies of the works

contained on the CD-ROM as part of its "library" package.

Although its texts contain no descriptive markup (e.g. SGML) , the books published on the Mayibuye Centre's CD-ROM *Apartheid and the History of the Struggle for Freedom in South Africa* represent the most successful achievement in a South African context of establishing literary web. The aim of the Centre is to "help recover areas of South African history which were neglected in the past and to create space for cultural creativity and expression in a way that promotes a process of change and reconstruction in a democratic South Africa." The SALIT project shares this aim, but aims to be more inclusive, extending its "library" to texts beyond those specifically related to South Africa's liberation struggle. Included on the Mayibuye CD-ROM are the full texts of fifty books published between 1978 and 1994, including biographies, poetry, and novels (e.g. Mary Benson's *South Africa: The Struggle for a Birthright* and Alex La Guma's *And a Threefold Cord*). Like Mayibuye, the SALIT project includes multimedia material in addition to the electronic texts. Electronic publishing lends itself to multimedia presentation and can integrate photographic, audio and film/video along with any texts it may contain as we have seen in the discussion of CD-ROM encyclopaedias (Chapter 3) where text, graphics and sound files are integrated into the entries. Turned to literary purposes, these features can provide an immensely rich resource, as the Mayibuye publication or Landow's *Victorian Web* evidence.

The Women Writers Project, in contrast to Landow's work in the "Victorian Web", has from the start committed itself to the encoding of its texts in SGML. Mah and Flanders (1996: 1) explain that SGML allows the texts to be encoded to help the user to navigate within the document, analyse it, compare it to others and find words and concepts within particular contexts (e.g. speeches, quotations, or notes). The TEI SGML encoding system for humanities texts also contains tags that allow parallel versions of the text: one with the original typographic and other errors (that may be important for research purposes) and

another more readable "corrected" text. The careful coding of texts has led to the Women Writers Project becoming an important scholarly resource in the field of pre-Victorian women's literature.

There is a tension between the more expansive "outward linking" hypermedia approach to electronic literary texts (typified by Landow, and the Mayibuye CD-ROM) and the more scholarly and systematic encoding approach (represented by the influential TEI/SGML project, and the implementation of its coding by the Women Writers Project, the Oxford English Dictionary and the Canterbury Tales Project). In the context of South African literature, there is a necessity to produce a usefully large corpus of representative texts as quickly as possible, which implies the former approach. With the limited resources available within the SALIT project, the opposing pressures of bibliographic comprehensiveness and detailed scanning and encoding of individual texts led to a decision to postpone the TEI markup of the texts in the virtual library to a later phase of the project. As a result, encoding was restricted only to those texts chosen for close textual analysis (Pringle's *African Sketches* and Mda's play *The Hill*) with text tagging conforming to the TACT program (See Chapters 6 and 7). To ensure the exchangeability of the SALIT texts with other researchers, they will at the very least have to be encoded in TEI format.

The SALIT Web and the virtual library it contains has the potential to become the structure on which a comprehensive digital library of South African texts could be based. The work described in this thesis is only a first step in the direction of a more widely collaborative project aimed at the collection, scanning and SGML/TEI markup of South African literary texts that could be used as shared "data bank" similar in scope and function to the Oxford Text Archive. The rapid development of electronic text collections around the world points to the building of a collection of local texts in electronic form as a research and development priority in South African literary studies. Large scale inter-textual studies in

South African texts are at present limited to print-based analysis achievable only by laborious manual methods, in which routine counting and matching robs the scholar of time better spent on interpretation. Compared to post-structural studies being conducted in already encoded corpora like the ARTFL in French literature and history, the Brown University Women Writers' Project or classical and medieval studies (the Thesaurus Linguae Graecae at Irvine) similar ventures in South African literature are severely inhibited by the time-consuming methods available. The more South African texts available in electronic form, the more feasible it becomes to conduct inter-textual research in the area. Of course, a South African electronic text resource, suitably encoded in TEI/SGML (see Chapter 6) could also then be analysed together with the rest of world literature.

Current projects aimed at constructing a South African Literary History would derive invaluable support in an endeavour which made early texts which may previously have been marginalised or suppressed freely available for research. It would also represent an inexpensive publishing alternative for texts which would otherwise be commercially non-viable. (There is nothing to stop any one right now scanning an out-of-copyright novel - *Mhudi*, for instance - and publishing it on the internet: Project Gutenberg is just such an example (see <http://jg.cso.uiuc.edu/pg_home.html>). Assuming that all publishing is likely to be primarily electronic within the next few years, it would be feasible both for public access and from a commercial point of view for publishers to collaborate in a project aimed at creating a comprehensive South African Literary E-text Centre/Resource. Book publishers must surely take heed of the current spectacle of newspaper managements hastily launching awkward electronic editions as their strategic planners predict the rapid shifting of advertising revenues from print to electronic media. Some major international publishers have already launched their own electronic catalogues.

The South African government is accelerating the provision of the infrastructure necessary

for electronic information exchange and has led the way in making draft legislation, for example, available on the Internet. The government site is itself a good example of a hypertext web in action - the Brown Postcolonial and Postimperial Web, as mentioned above, includes a link to a CIA information source on South Africa! Similarly, the *Mail and Guardian* on-line archive provides a valuable source of contemporary literature and film reviews. South African government Green papers and reports - for instance the *Report of the Commission on Higher Education* - become much more widely available as electronic texts - many readers do have access to printed copies. And even with rudimentary word-processor search facilities, a document like the *Report* is easier to search than the print version and was certainly cheaper to publish. The National Information Technology Forum (NITF) and the International Development Research Centre view the provision of internet access to rural and disadvantaged communities as a national (and African) priority. To some extent, widespread access to computer networks is already with us: in higher education, business and government. South Africa is currently ranked as fourteenth in the world in the number of World Wide Web servers in use, and the trend is towards the doubling of this capacity every year. While development in South Africa is being driven both by democratic notions of access to information leading to community empowerment, and by business interests, a concomitant of the expanding infrastructure is its capacity to disseminate electronic texts widely and inexpensively to its citizens. Given this trend, the SALIT Web could be accessed not only by the "meta-text" of a bibliography/encyclopaedia on CD-ROM, but also by other meta-environments designed specifically for teaching and learning. The implications for the SALIT Web for teaching and learning are explored in the next chapter. Not surprisingly, most of the international electronic text centres were either developed primarily as teaching resources, or have later been adapted to do so.

The essence of hypertext is the dynamic linking of any and all available material while at present CD-ROMs are protected by copyright law developed for books. For this reason, the texts in the SALIT Web are all outside the fifty year copyright threshold, apart from those

for which special consent will be obtained. The CD-ROM design, however, looks forward to a time when copyright laws are revised to resemble those that apply to recorded music or to cable television providing pay-per-view access to contemporary electronic texts over networks. (Jensen, 1992; Garson 1996; van der Merwe, 1996). However, right now the user is able to access thousands of texts in the public domain from Literature webs and virtual libraries on the Internet, and increasingly these are being encoded in standard formats that make the texts readily searchable with the tools available in the SALIT Web. However, very few of these texts are South African. Most of the large electronic text centres have adopted the principle of shared resources for scholarly purposes and have anticipated the lifting of present restrictions imposed by copyright and licensing. The Oxford Text Archive, for example, will provide individual modern texts to users who intend to use them for academic purposes only. In theory at least, electronic publishing has moved the book beyond its tangible, print-based form: as an electronic text, a book is not only easier to update (i.e. in the production of new editions) but has attributes that are fundamentally different from a printed text that are fully exploited in the Literature webs and in the SALIT Web. The challenges posed by existing copyright law are not minor considerations: the tension between the protection of individual intellectual property and public access yet to be resolved. However, the adoption of the international standard (SGML) by leading electronic publishers, notably Chadwyck-Healey, Oxford and Cambridge University Presses, and Macmillan has begun to lay the basis of some yet to be agreed sharing of copyright protected material, probably in the form of a pay-as-you-use system modelled on the music and television industries. The university-based electronic text centres already employ a similar principle, but on a less granular scale: the University of Virginia Electronic Text Centre, for instance, makes the copyright-protected parts of its collection available only to users with direct university affiliation. This material consists of texts or corpora that the University has bought outright or for which it pays licence fees, and the costs are recovered by student fees and other funding. Non-copyright public domain material is, however, made available to anyone with access to the internet, and these texts made be browsed via the

University's World Wide Web home page: <http://etext.lib.virginia.edu/uvaonline.html> .

While the SALIT Web Library option places the user and the texts together without any significant intervention from the web editor/designer (apart, of course, from the silent "canonisation" implied by the inclusion of those particular titles), the Period Tours are overtly expository in nature. Like a personalised rendering of the topical keywords feature, the tour guides the user through a "Literature Web" consisting of an illustrated account of the selected period, with hypertext links to associated full texts, reviews, photographs, maps, sounds, video clips and historical background relating to the time. Although the linear surface structure impels the reader from introduction to conclusion, the hypertext milieu offers multiple tangential explorations. Although the entire SALIT Web is a hypertext, the Period Tours take the most advantage of the non-sequential essence of the medium, offering the user a truly open-ended opportunity to read, pause, examine related texts in depth, return to the Tour or simply skim through. The Period Tours available on the current version of the CD-ROM consist of historical surveys of five main periods of South African literary history:

0000-1652 (from Time Immemorial to the arrival of Van Riebeeck)

1652-1795 (from Van Riebeeck to the first British occupation)

1795-1900 (British Occupation to the beginning of the 20th century)

1900-1948 (from beginning of the 20th century to the institution of apartheid)

1948-present (Apartheid and post-apartheid)

Under each period item there is a link to the illustrated tour itself; to significant dates of the period; to a dictionary of significant words related to the literature of the period; and to an index of significant words and various important discursive formations relating to the period. Currently the most developed Period Tour is the first one in the list, covering the

period from "Time Immemorial" to the first European colonisation.

Let us examine some of the features of this first tour "0000-1652 (from Time Immemorial to the arrival of Van Riebeeck)". The "tour" is a condensed narrative survey of the history of the period. Embedded in the tour text are hyperlinks to texts referred to, or to the topic under discussion, with important words and sections highlighted in red typeface. For example by clicking on "shamanism" next to *Clashing of the rocks*, the user is taken to an excerpt from Chapter 9 of Lindsay's study and clicking the "sound" icon, plays back a short recording of San speech. The text is further illustrated with pictures and segments of primary texts (indicated by a colour frame) and is sometimes accompanied by translations (see Marais' "Die dans van die reën"). The immediate accessibility of translations is a major contribution made possible by the hypertext environment to a multi-lingual literary resource like the SALIT Web. Mention is made in the next chapter of future plans to incorporate English translations of African language texts into the library, so as to broaden research opportunities into this neglected area of South African literary study.

As can be seen in the above illustration, it is not only the hypertext features that may be exploited fully in the tours, but multimedia elements as well. During the development of the SALIT Web, a small number of sound and video recordings were included for the purpose of experimenting with the full capabilities of the medium. The priority in the the first phase of the CD-ROM development was the inclusion of a comprehensive top-structure of bibliographic references, while scanning and preparation of the Library full texts was second in importance. The creation of an image library ranked third in the list of priorities and now consists of over 500 pictures, mostly photographs of authors. With only primitive equipment at our disposal, the process of finding and converting sound and video recordings to digital form has been postponed to later in the SALIT Web project development, although I have included the few sound and video images that were digitised

on the CD-ROM. However, one of the more extended multimedia experiments deserves mention here. Designed as a spoken tour, "The Shamanic Formation in Southern Africa" consists of a five-minute illustrated commentary on the shamanic form, and in particular the trance dances of the /Xam. The tour was produced in MS Powerpoint format and was originally hyperlinked to the "0000-1652 (from Time Immemorial to the arrival of Van Riebeeck)" Period Tour. Using images of authors, rock paintings and other illustrations from the image library, together with a recorded commentary from a script written by Johan van Wyk, the Tour "talked" the user through the subject area in the equivalent of a short film or tape-slide sequence. Although it played as one seamless presentation, the tour consisted of 35 images (some of them animated) and eleven sound files. However, because the tour was likely to draw too heavily on the computer memory and processing resources of likely users, it was excluded from this version of the SALIT Web.

Images and commentary from the first part of the presentation are included here by way of illustration (Figure 16):
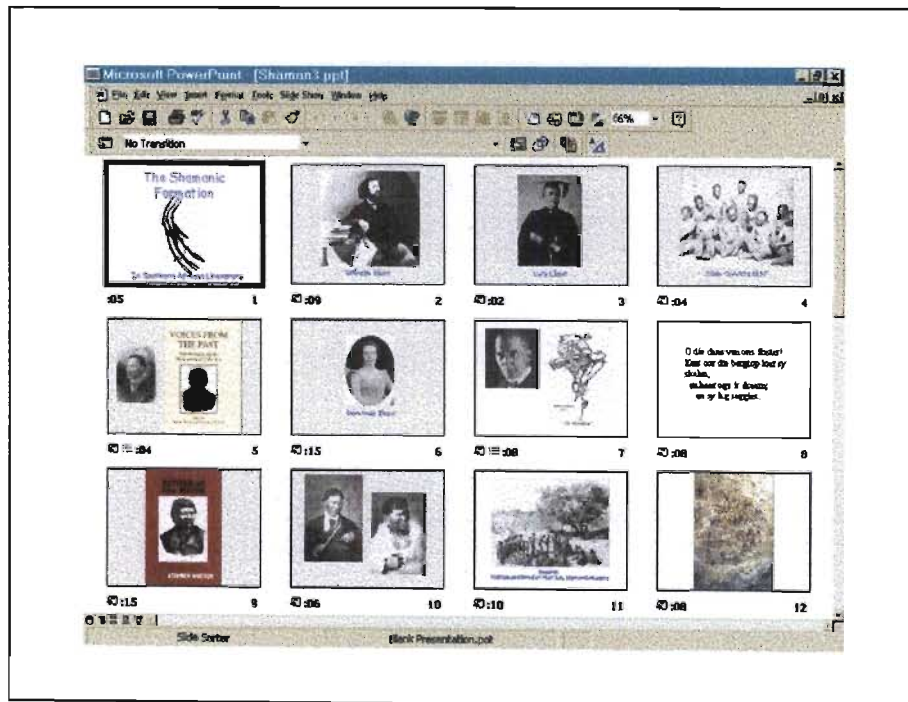
Figure 16: Opening Slides in "Shamanic Formation" Tour

Shot no.	Picture	Commentary/sound
1	Rock painting 01	San speech clip from "The day the white hunters hunted us"
2	Bleek photograph	The poetry, songs and stories of the hunter-gatherers were recorded for the first time by Dr Wilhelm Bleek, a German linguist.
3	Lucy Lloyd photograph	and his sister-in-law, Lucy Lloyd, in 1857.
4	/Xam convicts 1837 photograph	Bleek took some /Xam convicts into his home in order to study what he called "their mind"
5	Voices from the Past book cover/ San photo 15	A number of other recordings have followed since then ...
6	Dorothea Bleek photograph	
7	Marais photograph (Slide from right) "Outa Hendrik" drawing	Another important collection of hunter-gatherer lore, Eugène Marais' *Dwaalstories* appeared in 1927. These stories were told to Marais by a farm labourer "Outa Hendrik".
8	"Dans van die reen" extract	Integrated into the stories are poems such as "Die Dans van die Reën" which has become a classic in Afrikaans literature...
9, etc.		

(Van Wyk and Stewart 1997)

"The Shamanic Formation in Southern African Literature" tour revealed the considerable scope that exists within the SALIT Web for the development of multimedia material. A solution to the digital format problems is available in the form of "streaming" applications like RealVideo, which we adopted for the sound clips on the CD-ROM. Only one session of

original recording was conducted during the project: South African poet Dennis Brutus (Figure 17) agreed to read a selection of his own poems and those of Arthur Nortjé for videotaping. One reading from this session "Silence in the still, warm room ..." was digitised and included on the CD-ROM. Additional sound and video clips include readings by other contemporary South African poets, and an extract from a filmed performance by a praise singer.

Figure 17: Dennis Brutus video clip

Literature webs are used for both research and in learning/teaching contexts, representing the convergence of database - encyclopaedia - archive into a dynamic resource that de-centres the compiler/editor/teacher unique to hypertext. The learning and teaching possibilities based on electronic resources like the SALIT Web, are explored in the next chapter.

Chapter 5 Learning

Development of the computer as a teaching and learning aid has accompanied the general development of the micro computer since the 1970s. To begin with, teaching programmes (Computer Based Training, CBT) focused on programming - how to make your computer calculate a formula, or list your personal library catalogue - and later on programmed "teach-test-teach" packages that presented information to be learned, followed by multiple choice or true/false quizzes that determined whether the student should go back and learn again, or proceed to the next lesson. Many CBT programmes still follow this latter pattern, and for some disciplines such as physiology or second-language learning are entirely adequate - providing the student with a form of individual tuition that can accommodate different rates of learning. For literary studies, however, such rote learning programmes are usually inadequate except as simple progress tests.

Since the advent of hypertext, a very different set of possibilities has arisen for the architecture of computerised learning programmes. As is clear from the discussion of a resource like Landow's "Victorian Web" (see the previous chapter) hypertext offers the capability of discovery, rather than a linear progression through pre-determined facts. Given a sufficiently rich resource base, a student could access primary and secondary texts (together with graphic and video material) in order to formulate answers to questions posed by their tutors or by specially prepared computer guides. The student's progression from one source to another can be automatically stored, leaving a trail of links that constitute an electronic bibliography that may be revisited by subsequent learners.

Any consideration of the role of computers or hypertext in literary studies, however, raises the issue of approaches to learning, what Papert (1986) identifies as the

> ... central question for educators ... whether schools of the future will go on teaching the
> same curriculum, using computers to do the job better, or whether we'll see radical change
> in what is taught and what is learned in schools (Papert 1986 xxxv).

It seems likely that the assumption made by Cumming and Sinclair (1991: 316) that
hypertext works against transmission (rote) learning models would be apparent in the design
of most learning materials in this medium. Their findings do support this conclusion, as they
found that none of the programs that formed part of their investigation "conceive[d] of
learning in the rudimentary modes of drill-and-practice or simple skills rehearsal".
Grapevine, one of the programs they review, resembles to some extent the Victorian Web.
Grapevine consists of over fifty-four works covering the socio-political history of the 1930s
as a context to a study of Steinbeck's *Grapes of Wrath*. The linked material consisted of
books, film clips, photographs, audio tapes, magazine and newspaper articles. Information
relating to the social context of the novel was indexed under thirty-three topics, enabling
over a thousand links between the novel, the materials, and topics like "dust bowl" and
"alien labor". It was planned to add facilities like on-screen notepads and authoring
programs to make the materials more interactive: facilities that are provided in both the
electronic projects in Folio Views that supplement this dissertation, and in Asynchronous
Learning Networks (ALNs - see below).

The most recent developments in on-line teaching and learning are variously called
"Asynchronous Learning Networks" or "Distributed Learning", combining several
components of commonly-used on-line communication software applications already
familiar to computer users in academic institutions - e.g. e-mail, internet web pages and
discussion groups - to provide a holistic teaching and learning "environment" to bridge the
often insurmountable technical problems that bedevil the separate and often incompatible
collection of programmes on university networks. Shared network communication software

such as Lotus Notes or Microsoft Exchange were developed as an answer to this very difficulty identified by organisations as impeding collaborative information-sharing. The adoption of "learning organisation" philosophies (e.g. Senge et al.1994) in large organisations - private companies and academic institutions alike - has promoted the recent development of coherent ALN packages. Learning organisations require the development of collaborative teaching and learning "environments" rather than linear programmed learning, or inaccessible databases of discrete information fragments. Murdoch University's Teaching and Learning Working Party (1998) identified thirteen software packages in their review of currently available resources. Amongst these were ClassNet, LearnLinc, Learning Space (discussed more fully later) and the Socrates Program. A similar comparative exercise can be found in Kerry McCollum's article "Colleges sort through vast store of tools for designing web courses." (McCollum 1997). McCollum considers the pros and cons of choosing a commercial product, rather than one developed as part of a university research programme. Both types of program have their supporters but, as will be seen from the discussion that follows, for literary studies, where ease of access is of prime importance, a "ready-made" commercial package is the most suitable.

Both program types are represented in the three successful implementations of ALNs to be discussed here: the University of Pretoria course in Computer-based Communication and Distance Learning; the University of Potchefstoom's MBA programme (in collaboration with the Africa Growth Network); and the University of Greenwich's Post Compulsory Education and Training (PCET) courses. All use a model of asynchronous learner access that would suit a learning application of the SALIT Web.

The difference between "synchronous" and "asynchronous" networks lies primarily in the time the user would have to be connected to the Internet. Where synchronous activity requires on-line connectivity to make use of the latest "real-time" Internet plug-ins that

support synchronous chat and streaming ("live") video, asynchronous activity allows the user/learner to read material and complete assignments off-line in his or her own time, and requires only brief Internet time to use e-mail to send and receive messages and assignments. An ALN is therefore also more economical for the user, providing a rich resource base on the home computer, but with seamless access to the more interactive components of the course when necessary. Malcolm Ryan of the University of Greenwich cites further educational advantages for the learner:

> ... there is evidence to suggest that asynchronous technologies are better able to facilitate the reflective models ... embraced by many CPD [continuing professional development] programmes. Additionally, the time-location independent nature of asynchronous media ... may facilitate greater flexibility and student autonomy. (Ryan and Culwick 1997: 1)

These implementations offer studies in the humanities, and literary studies in particular, effective solutions in the form of learning environments that combine flexibility and maximum learner participation with the minimum of technical computer competence on the part of either learners or instructors. Katchoff and Ryan (1997) found evidence that use of an ALN helped users overcome the "technophobia" often associated with computer use. The latter two courses (those of Potchefstroom and Greenwich) use the Lotus Notes/ Lotus LearningSpace packages that would provide ideal environments for literary studies. By contrast, the aim of the Pretoria University course - training educators in the development of ALNs - is helpful in underscoring the need for seamless, uncomplicated media for computer-based learning and in identifying the principles and strengths of networked learning.

Lotus Institute, the research and development division of the company found that

> Collaborative computing research ... showed that the highest level of learning occurs in an instructor-facilitated, team-centered environment with anytime, anywhere access to rich

content. As a result, LearningSpace modules were designed to encourage all aspects of such a collaborative learning environment. The modules allow users to engage in problem-solving activities, debates, discussions and exercises that result in the creation of new knowledge; manage and leverage knowledge that is key to an organization's success; access and use stored information; and receive personalized feedback from instructors. (Lotus Institute 1998)

Behind the corporate-speak of this announcement, is a sound educational principle that answers the needs of the South African learner in the context of on-line literary studies. If we were to rephrase the Lotus Institute in terms of a learner-centred approach to the study of South African literary texts on-line, we might come up with the following: The development of literary studies in South African literature depends on the provision of an instructor-facilitated, team-centred environment with anytime, anywhere access to rich content such as that contained in the SALIT Web, for example. A collaborative learning environment of this type would allow students to explore literary critical assumptions, engage in debates and discussions with instructors and fellow students, and complete exercises that result in the creation of new knowledge; consult and use existing critical and creative writing to guide their own research; access and use stored information, texts and background material; and receive personalised feedback from instructors.

In their "White Paper: Distributed Learning" (1998), the Lotus Institute pinpoint "technology limitations" (p.3) as having impeded the use of ALNs in situations requiring collaborative learning. They put forward three broad outcomes required by an educational programme:

- information transfer
- skills acquisition, and
- mental model change

While the first two outcomes are largely self-evident, the third is less straightforward, yet most congruent with the kind of learning required within the context of current South African education trends, viz. the competency-based, learner-centred requirements of the National Qualifications Framework (NQF). Let us briefly review some of the key elements of the NQF to examine the close relationship between the principles and architecture of the SALIT Web and the new South African education system. The 1996 *Discussion Document: Lifelong Learning through a National Qualifications Framework* (prepared by the then Committee for Development Work on the NQF, now subsumed into SAQA), highlighted the following conditions to be met for the realisation of the "lifelong learning" advanced in the *White Paper on Education and Training* (1996):

> * Alternative learning programmes must exist for a range of learners who are unable to attend education and training institutions either on a full-time or on a regular part-time basis.
> * Learners must be equipped to take advantage of open learning and multi-media education and training possibilities. (1996: 21)

The open learning opportunities inherent in the application of the SALIT Web to teaching and learning are of special relevance here. Further, ALNs are designed to support what the *Discussion Document* calls "changing theories of language, learning and cognition which, broadly summarised, could be said to be moving away from a "transmission" model". As has been noted above, ALNs actively promote learner "participation in creating and enjoying knowledge" (1996: 28).

The "mental model change" referred to by the Lotus Institute is brought about by learner participation in the creation of new knowledge, in particular through the experiences of a team, or collaborative learning. The term "mental model" refers to Peter Senge's influential work in the field of organisation learning (Senge et al. 1994) and has been described as a set of "deeply ingrained assumptions, generalizations, or even pictures or images that influence

how we understand the world and how we take action" (Lotus Institute 1998).

Discovery is at the heart of the learner-centred approach, and in pedagogical terms challenges the primacy of rote learning or the "transmission" model referred to in the NQF, by encouraging learners to internalise knowledge through observation and experience, rather than by simply reproducing facts. The role of the teacher in the collaborative model shifts from "transmitter" to a facilitator, who collaborates with learners to design individual learning experiences that in turn lead to students' developing the skills and resourcefulness to create their own knowledge.

Here we return to the successful implementation of ALNs at Pretoria, Potchefstroom and Greenwich universities. At the University of Pretoria, Johannes Cronjé's research (Cronjé 1998) aimed to "determine the feasibility of using the Internet for co-operative, constructivist, distance learning. The following research questions were asked:

> * Can a "virtual classroom" be successfully simulated on the Internet?
> * Will students who have been used to contact teaching adapt to the virtual classroom?
> * To what extent could co-operative work be done over the internet?
> * To what extent will technology and technological illiteracy affect the learning process?"
> (Cronjé 1998: 1-2)

For the purposes of literary studies, Cronjé's findings suggest that although the broad principle of using an ALN to develop literary skills and knowledge is highly appropriate, the level of technical expertise required to learn and use an array of existing on-line communication packages is beyond the scope of what a literary course would hope to achieve. A "groupware" software package, like those used at Potchefstroom and Greenwich overcomes the latter obstacle. Ryan and Culwick (1997) describe groupware as "essentially any computer application that runs over a network, and/or is accessed by direct dial means,

allowing the user group to increase their communication, collaboration and co-ordination."

Among the positive benefits of the ALN, Cronjé mentions the success of the medium to support co-operative learning, and to promote on-line debate via an e-mail-based discussion group (listserv) in particular where this is linked to a wider forum. He used *Itforum* to good effect in his study, and the SALIT discussion group of the CSSALL could (and is) playing an important role in encouraging local and international debate and information-sharing in the field of South African literary studies (see Chapter 9). The use of an Internet website as a locus for his "virtual classroom" had additional value:

> A web site is a good place to store large quantities of information for non-interactive
> purposes, while creating ownership by allowing students to add their own URLs
> contributes to the interactivity of a usually passive site (Cronjé 1998: 7).

The SALIT Web is designed to underpin just such a repository of "large quantities of information" with similar implications for wider use, including the interactivity described here.

Of course, where the aims of Cronjé's project, and that of the teaching and learning application of the SALIT Web diverge, is in their subject matter. The Pretoria course forms part of a Masters Degree in Computer-Assisted Education, so students need to learn ALN development skills as an explicit requirement. The problems they encountered with incompatible formats and the like provided the learning experience they needed to achieve the desired outcome of the course: mastery of the instructional medium. The outcome of a literature course is the meaningful interpretation of literary texts, so we would wish to minimise, if not entirely eliminate, technical difficulties. Hence, the commercial groupware package option used by the Potchefstoom and Greenwich universities is more suited to the needs of this study.

Chris Derby describes an implementation of Lotus LearningSpace in delivering a distance education MBA at the University of Potchefstroom (Derby 1998). He was involved with the course as a commercial service provider (through the company Africa Growth Network) and as such was not involved in any way with the *content* of the course, yet closely involved in the instructional design and delivery of the ALN. The model adopted for the course design was a collaborative, open learning system and included the provision of on-line full-text articles and other text-based course material, recorded video (e.g. interviews with business leaders), conferencing (live satellite video conferencing), e-mail discussion forums (both lecturer-to-student and student-to-student), on-line self-evaluation (multiple choice type tests) and the submission, marking and return of marked assignments on-line. Amongst the advantages of using the Lotus LearningSpace platform, was the capability of the system to ensure that identical copies of the learning resources were downloaded to all learners, including the automatic updating and management of computer files, so that the misplacement or haphazard indexing of files by inexperienced computer users could be avoided. Because the bulk of the LearningSpace resources resided on the users' own computer hard-drives, most of the student interaction with the material occurred off-line, requiring an Internet connection only during the sending and receiving of e-mail and during the downloading of update material. Incompatibility problems like those experienced by Cronjés students were circumvented by the common "sharing" of one e-mail interface, and even in the case of submitting assignments, LearningSpace allows "cutting and pasting" of word-processed material into its own pages so students could use any one of the multiplicity of word-processing packages to compose their work, without having to undertake any conversion for submission of the material to their tutors. LearningSpace was specifically developed for a university teaching and learning environment, and the interface includes:

- a "schedule" which outlines the process of the course of study

- a "media centre" including short video segments and CBT courses (the self-evaluation tests referred to above)
- a "course room" section where the main learning activities are situated - including e-mail,
student notes and full-text resources
- a "profile" in which a photograph and personal information on each student was available
- an "assessment" database (linked to each student, but available only to lecturer)

LearningSpace thus provides a shared, "virtual classroom" which can be accessed on-line for same-time interaction, or in an asynchronous, disconnected manner. The application of this type of software in literary studies would allow learners to explore the full texts in the database and navigate to information based on their own interests as well as their levels of competency. The *African Sketches* text, for example, is linked to articles on literary history and theory, but also to visual material and maps depicting the physical environment of the Eastern Cape of the early 19th Century. The *Survey* provides an authoritative, but accessible and visually appealing introduction to South African literature, while linking to serious scholarly research that learners would be able to use later in their course of study. Student assignments represent new knowledge that is stored in the learning environment, thus bringing the learner directly into the process of augmenting and building a richer archive. What follows here is an adaptation of The Lotus Institute's *White Paper* simulation of the learner's experience in entering an ALN course for the first time. We will follow the experiences of a fictionalised learner, Mandora Mdlala, a student in that course.

Mandora works full time and chooses to participate in a distributed course because he cannot afford to take time off work to attend classes at a scheduled time and place. He lives too far away from the university to travel to regular part-time evening classes.

Mandora is beginning a new course in LearningSpace being offered by a university several hours away. **(1a)** He starts in the course Schedule (Figure 18) by playing the instructor's video clip of introduction materials, including a brief "visual tour" through the history of South African Literature drawn from photographs, maps and illustrations from the SALIT Web. He then browses the Schedule module to gain a complete understanding of the course learning objectives and expectations. One of the first entries in the Schedule directs Mandora to the Profiles module where he is instructed to edit his profile.
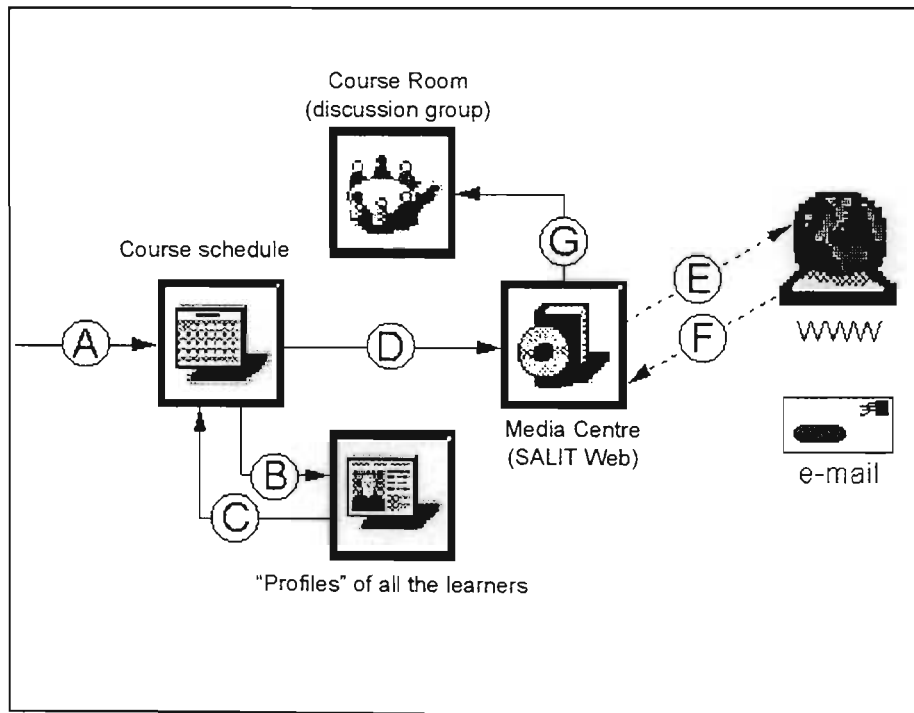
Figure 18: Student interaction in an ALN (1)

(1b). Mandora navigates to the Profiles module and finds his name under the participant category. When he opens his profile, he clicks on the "Edit Your Profile" button. From there he enters his contact information, education, work experience, and his interests. He also adds a picture of himself, writes about his interests and adds a short poem he has

composed.

(1c) Mandora navigates back to the Schedule module, goes to the Personal Progress view and marks this task complete. He subsequently moves to the next document, which instructs him to read an article on Oral Literature and watch a Powerpoint presentation in the MediaCenter on the Shamanic Formation in Southern African Literature. (1d) He links to the MediaCenter and completes his assignment. While in the MediaCenter, Mandora would like to learn more about Oral Literature and in particular, praise poetry, identified there as a specific topic area. (1e) He clicks on the WWW SmartIcon and accesses the World Wide Web for additional information. While browsing on the Web, Mandora finds the Commonwealth Literature Library's list of resources on "Oral Literature From Africa and The Caribbean" that includes a reference to *Foundations in Southern African oral literature* edited by Russell H. Kaschula, (http://www.commonwealth.org.uk/ readlists/oral.html) information that he thinks his fellow students would also be interested in reading. He copies the first page of the Web site including the hot links, and asks if it is possible to access the book on-line. (1f) He goes back into the MediaCenter, navigates to the CourseRoom, (1g) and "Starts a Discussion" on the subject of praise poetry and its influence on current South African poetry. Mandora pastes in the captured Web page to share the information in the CourseRoom with his fellow students. Mandora "leaves" the LearningSpace environment and plans to complete the next exercise the following day when he has a day off from work.

(2a) At the Community library resource centre (Figure 19) near his home, Mandora uses one of the Internet linked computers to re-enter the Schedule where he left his last completed task. The next assignment is a group project with a team that has been established by the instructor, involving the collection of contemporary accounts of the British colonisation of the Eastern Cape. He reads the objectives of the group project and finds the name of the team in which he will participate. He notices that the "View Related

Discussion" button appears at the top of the screen indicating that a discussion about this document exists in the CourseRoom. **(2b)** He clicks on the button to view the current discussion.
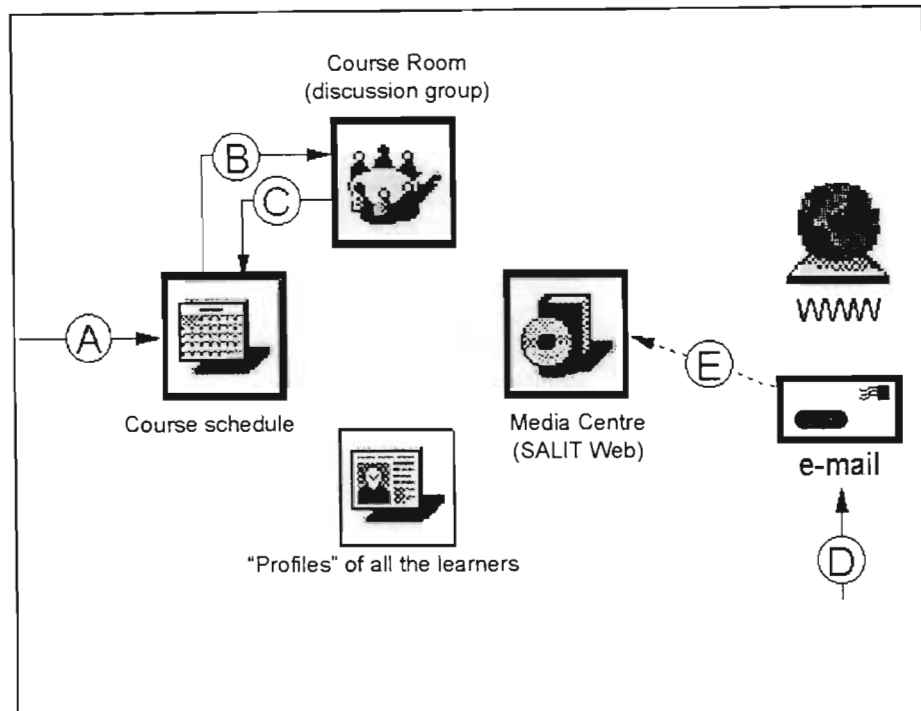
Figure 19: Student interaction in an ALN (2)

He finds several documents, one of which was entered by a fellow team member. He opens the document and reads his teammate's ideas. He clicks on the "Comment" button and adds his suggestions and ideas. He also notices a document that was entered by the instructor, which he reads. Mandora has several questions and creates a comment on the instructor's document requesting a response to his question. The instructor also suggests that he subscribe to the SALIT discussion group, where there has been an ongoing debate on the nature and significance of Oral Poetry in South African literary studies. Before leaving the CourseRoom, Mandora decides to read some of the other discussions that are taking place within the team as well as those for all students, and adds a few more comments to these discussions. One of the resources listed for the task is Pringle's *African Sketches* and

Mandora links to the full text provided with the original LearningSpace download. Within the document, he links to the poem "Afar in the Desert" in which the course material has suggested there is a reflection of Pringle's ambiguous feelings towards the country in which he has made his home. A hotlink plays an audio recording of someone reading "Afar in the Desert".

(2c) Mandora then re-enters the Schedule module to complete his first quiz. He opens the document and clicks on the "Start Here" button to begin. He completes the questions, which include true-false, multiple choice and several short answer questions on the overview of South African literature provided in the SALIT Web Period Tours (see Chapter 4) and clicks on the "Submit to Instructor" button. When Mandora returns to the Community Centre the following day, he connects and sends his additions to the server and receives additional information entered by other members of the course.

A few days later Mandora's quiz is returned via e-mail. (2d) The returned, graded quiz displays the correct answers, Mandora's answers and an explanation of the differences between the two. He also receives an annotation from the instructor about areas in which he excelled and those in which he may want to spend more time. The instructor has included a link to content in the MediaCenter, and following this link, (2e) Mandora begins his work in the next segment of the course.

Mandora's fictional experience described here is speculative only in terms of content. Students enrolled for the three courses considered in this chapter are participating in exactly the same type of learning environment, but with business studies or education as their course content.

The need for remote access for working teachers and trainers motivated the University of

Greenwich's School of Post Compulsory Education and Training (PCET) to find an ALN that would deliver some of its course modules. Their Certificate in Education Certificate in Teaching Competence is designed to provide initial teacher training to those who are already in employment in the public or private sector as teachers, lecturers or trainers. It aims to enhance their performance by helping them to develop teaching skills "whilst understanding the principles which govern the way learners learn and teaching organisations work".

Concept of CMC
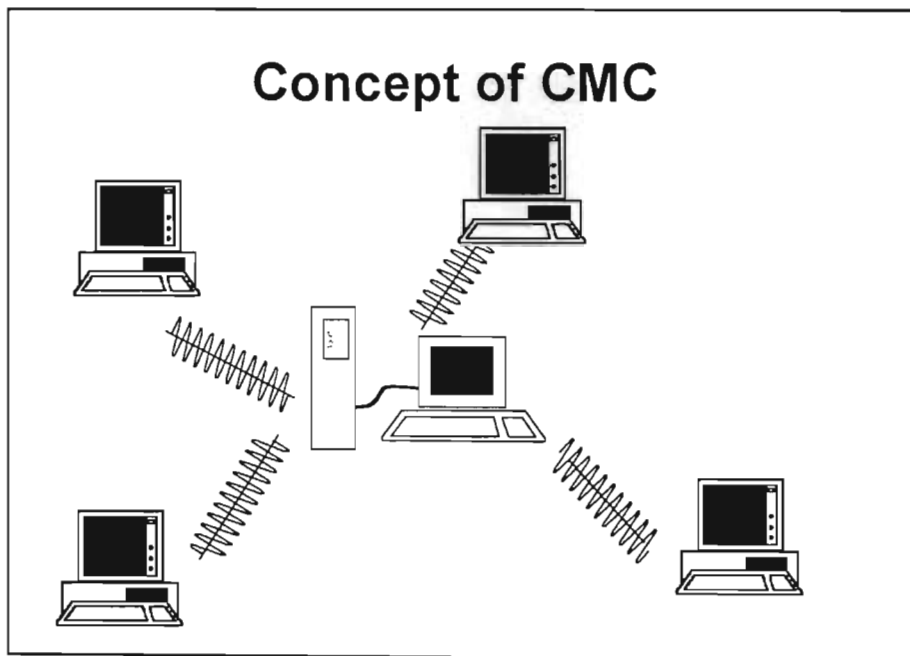
Figure 20: Greenwich's Computer Mediated Communication Model

The School has developed distance learning materials to support distance learners, and these have been in use since September 1994. As in the Potchefstroom project, Greenwich's VILE (Virtual Integrated Learning Environment) consists of

classrooms where a range of activities and discussions occur, an electronic resources centre to which both staff and students contribute, a tutor-free student common room for

social chat and virtual coffee, a staff room for on-going curriculum development meetings, a mail room and an advice centre. (Ryan and Culwick 1997: 2)

The similarities between the Greenwich VILE (see Figure 20) and the hypothetical South African Literary Studies learning environment experienced by Mandora Mdlala given above are clear. Katchoff and Ryan observed that the archived discussion of the learners in the groupware environment produced a valued resource: "The richness of the archived exchanges between participants aids the reflective process and students gain considerably from access to a team of staff who are able to provide different perspectives on issues and problems." (1997: 2).

What is the relationship between ALNs and the hypertext literary resources (*The South African Literary Encyclopaedia*, Full-text database, and the *Survey of South African Literature*) that underlie them? While ALNs help create a stable learning environment that encourages exploration of the texts, discussion and collaborative project work, the hypertext resources themselves encourage a much more flexible use of the materials. The hypertexts retain the "authority" of the original works, both creative and critical, yet open up new avenues of literary activity.

Geoffrey Rockwell, in his "Hypertext for the humanities" paper (Rockwell 1996), explains and illustrates how scholarly projects might draw from advances in audio, video and graphics capabilities now available within hypermedia packages, yet still retain standards of academic rigour, content and credibility. While highlighting the research and learning possibilities of hypermedia, however, Rockwell stresses the enduring strengths of conventional text, and in particular, the portability and ease of access of the book as compared to a computer. The strength of hypertext lies primarily in its non-sequential organisation of material and its potential for student learning through a process of discovery. A tutor in the context of an ALN would provide guidance and access points, as

in the "Mandora" example, but once within the hypertext environment, the medium has a uniquely expansive nature of its own.

The teaching and learning applications of the *Encyclopaedia* project aim firstly at providing a rich source of South African literary historical information, and secondly at making electronic versions of specific texts available for literary analysis all on the same CD-ROM. Rockwell gives the example of the hypermedia *Freud* archive to illustrate the ease with which students were able to access primary and secondary sources within the programme, and then create their own "webs" of links between portions of the material. Because such webs are separate from the text itself, it is possible for these individual "trails" (reminiscent of the associative trails postulated by Vannevar Bush (Bush 1945: 15) to be saved and shared with other users of the material. Rockwell suggests that an associative trail is to some extent an *argument* that can be evaluated by a course instructor. The "trails" created by individual learners or in collaborative projects could be submitted for assessment within an ALN, while simultaneously added to the "rich resource" noted by Katchoff and Ryan in their Greenwich VILE. As an introduction to the material, an instructor might at first encourage a student to follow a particular pre-designed trail, and then evaluate a student's own trails and annotations to the links.

The principle of individual interaction with the material is another central design requirement of the *Encyclopaedia* project. In the early planning stages of the electronic version of the *Encyclopaedia* it was considered vital that both researchers and students should be able not only to draw on the material as a resource, but create and develop the material to suit their own needs as they went along. Rockwell's analysis of the capabilities of several proprietary software "browser" programmes reinforces the view that the software adopted for our own project should provide a "web management" capability in addition to conventional layout and retrieval facilities.

Folio Views, used in the SALIT project, has features similar to the SoftQuad Explorer and
Macromind Director, both of which offer powerful hypertext features and supported SGML
encoded documents. Explorer uses an interface similar to DynaText, another prominent
browser programme (used by Cambridge University Press for their 1996 *Johnson, Chaucer*
and *Shakespeare* CD-ROM titles). Both programmes present an interactive table of
contents in a window on the left of the screen, and text on the right. Links are built into the
primary text leading to commentary in secondary sources, but the user may add her own
links, and commentary (as described above) and these webs can be saved for subsequent use
(or for evaluation by the tutor). Where Explorer and DynaText used the book as a metaphor
for their hypertext environments, by contrast The *Musée d'Orsay* CD-ROM provides an
entirely graphical interface into a hypertext environment - using real or imaginary space as a
metaphor for the information. Another project that used a graphical interface is the Piero
Art History Project, where the user's access to the database of linked historical, sociological
and artistic information is directed through three-dimensional images of Piero's frescoes.
Although a graphical interface was not considered the most appropriate for the
Encyclopaedia project, the possibility at some future stage of adopting this approach
(especially for users with weak literary skills) has reinforced the principle of
interoperability, or compatibility with varying types of software (see Section 2) as a key
design criterion for the present project.

Greenwich Case Study

The School of Post Compulsory Education and Training (PCET) at the University of
Greenwich conducted a study into the use of an ALN to deliver its distance course, the Post
Graduate Certificate in Education (PGCE). Using the learning environment described
above, the groupware Lotus Notes, the researchers/tutors identified a sample group of thirty

students by calling for volunteers from the registered intake of 250. Twenty of these students completed the course, but only two of the drop-outs were attributable to difficulties with the medium (insurmountable technical problems). Their findings in respect of learner reactions to the various type of electronic communication available in the course have a bearing on the design of a similar learning environment for the SALIT Web. Students preferred one-to-one contact with tutors (via e-mail) to the many-to-many conference features, suggesting that the nature of the learning material influences the type of interaction required.

> ... it was generally felt that the competence-based and reflective units - which required the compilation of portfolios rather than the production of essay type assignments - were less suitable for discussion between learners than the other more overtly content-based and discursive units. The presence of printed distance-learning support materials to guide an individual student in the preparation and production of evidence for a portfolio led to a need for consultation with the tutor rather than for a debate with peers (Lewis, Gould, and Ryan 1997: 6).

Where these findings have implications for the SALIT Web is in the importance of identifying learning activities that would be suitable both for literature study and for the level of the target population of students. In other words, those areas of literature study that are primarily content-based, such as literary history would require a learning environment with different features from those that required the application of critical skills. A course for undergraduates would tend to contain more modules of the former type, while a postgraduate course would expect considerably more learner independence and require more collaborative work and more peer group discussion amongst the participants.

Another case study with important implications for the SALIT Web, is that of the Learning Page of the Library of Congress. Researched and designed with the explicit aim of drawing secondary school learners into using the digitised primary sources available from the Library of Congress, several aspects of their development process and the resulting learning

environment recommend themselves to the South African project. Underlying the design of the Learning Page are the "clear, orderly analysis of the learner; a comprehensive analysis of the content to be learned and its organisation and presentation; and an analysis of the tasks and skills involved in the learning event" (Graves 1997: 677). Graves' analysis of the target audience for the Learning Page acknowledges the current shift in learning styles from a transmission model to one in which critical thinking and learner-centeredness predominate. She suggests that the critical skills of analysis, evaluation and synthesis of information are best developed by providing learners with primary source materials and setting project-based learning tasks. By substituting "The SALIT Web" for the American Memory collections contained in the Library of Congress useful lessons can be drawn from her Learning Page design guidelines that emphasise meaningful communication, data organisation and hypermedia characteristics. With the underlying aim of promoting research skills, the Learning Page seeks at one level to "chunk" content into meaningful groupings, highlight the interrelatedness of groupings by linking them, and then provide hot links (hypertext links) between these "pathways" and the database of primary materials. The equivalent of these pathways in the opening screen of the SALIT Web would be grouped resources labelled "Genres" (see Chapter 2), as well as the headings used in the Learning Page: "Topics" (including themes, motifs, and specific discursive formations), "Time" (chronological - timeline - listings of selected publications and events), "People", "Places" and "Events". The user would, of course, also be able to access the database in a more open-ended way, using a wide-range search tool across all the Database fields. Both Graves and Robert Janus, in his account of the conversion of the *World Book Encyclopedia* from print-based to digital form (Janus 1997: 607), stress the importance of providing both structured and unstructured search strategies. Graves calls the former process "searching" and the latter "browsing"; providing both types of search would significantly enhance the access to the SALIT Web.

Learner guidance in the form of "Pathways" or overviews are not new to encyclopaedia design: in the SALIT Web project they transform what is essentially an enhanced bibliography into an encyclopaedia. This is true of the Folio Views interface to the Database we call *The South African Literary Encyclopaedia*. The Learning Environment interface takes the "guidance" several steps further, not only guiding but actively propelling the learner into research within a "context rich" framework. The Learning Page of the Library of Congress includes a "Become an Historical Detective" activity that could be adapted for use within the SALIT Web Learning Environment:
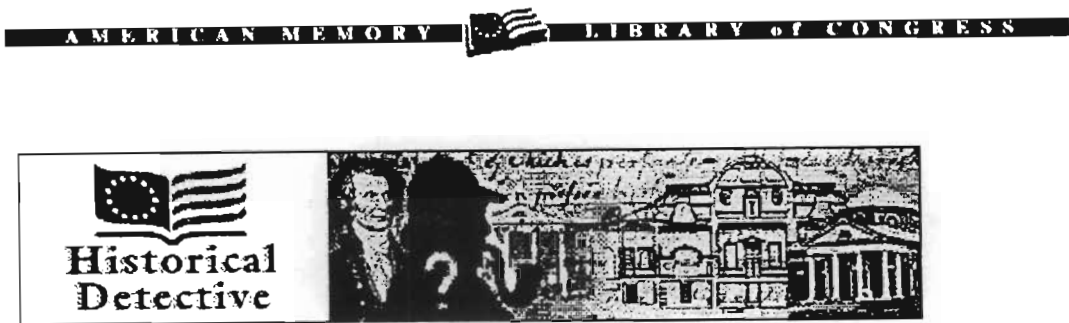
Become a Historical Detective
Search American Memory collections to test your D.Q. (Detective Quotient)!
Solve this mystery: *Who is this woman?*
The mystery woman says:
"Read my poem. Do you see autumn as I do? I wrote this poem when I was thirteen for my friend, Mr. A. G. Bell. Who am I? What colors do I see?"

Start | Clues | Solution

Start:
 Solving a riddle or puzzle is the same whether you are doing your detective work on the eighteenth century or the twentieth century.

- You must ask questions (who? what? when? where? why? how?), hunt for clues, talk to witnesses and visit the scene to search for evidence.
- You must form a hypothesis (I think ...... because ....) and gather evidence to prove your hypothesis.
- Your evidence must be authentic, first-hand information that you have carefully reviewed to make certain that it is genuine and will prove your hypothesis.
- Occasionally your investigation will uncover vital facts that will make you rethink your original hypothesis and make the necessary changes to solve your riddle or prove your case.

Clues
- What are your keywords? (Remember that the Library's search tool can search on subject terms and words in the photograph's caption in the bibliographic record of the item.)
- What Pathfinder Page(s) would have these key words?
- What collections have these keywords? The list of collections provides a brief description of the contents of each collection. This may aid your search.
- Brainstorm words. Use the keyword search tool to find evidence. Check your spelling (remember - YOU are the detective, not the computer).
- At a dead end? Use the alphabetical subject listing in each collection to give you some ideas. Scan the list to see what kinds of clues you can find.
- Try your keywords in the search box of the American Memory Collection Search Page or on the search page of individual collections.
- Need help? Ask your school media specialist or teacher. You can also send email to our Reference Librarian.
- Gather evidence. What is your hypothesis?
- Was your hypothesis complete? Did you find other forms of evidence that caused you to restate your original hypothesis?

Solution
If you are really ready to check your hypothesis, you can find the answer to the mystery here.

(The answer is: Helen Keller. She wrote the poem 'Autumn' at the age of thirteen, in 1893, for her friend and mentor, Alexander Graham Bell, inventor of the telephone and a teacher of the deaf. Search on *autumn*, poem to "Autumn," poem by Helen Keller, 27 October 1893 in *Words and Deeds in American History*.)

As in the Library of Congress Learning Page, the SALIT Web Learning Environment should also provide its users with "scaffolding" to support learners in their assignments and tasks, and to develop their evaluative and interpretive skills.

It must be stressed however, that the interactivity inherent in ALNs and in the type of Learning Environment described above are essential if we are to avoid the danger of multimedia in literature being reduced to passive entertainment as can often be the case in commercial applications. Rosanne G. Potter, Professor of English at Iowa State University warns of the "seductive lure" of hypertext "tempt[ing] teachers into becoming mere

facilitators and turn[ing] students into spectators" (Potter 1996: 182-183). Instead of promoting learner-centred activity, literary studies could be confronted by a fate worse than the "transmission model"- the Hollywood model. Potter is a well-established researcher and commentator on the use of electronic texts in literary studies, having taught and published on computers in literature since 1985. However, she counterbalances her concern about students "clicking through" and "sliding irresponsibly" over the surface of literature by citing examples of analytic or creative computer-based activities that engage students with the text, or enable them to participate in meaningful discussion. Here, the ALN capacity for conferencing via e-mail (MUDs and MOOs[5]) or "computer-mediated discussions" is affirmed as a powerful learning medium, especially in the literature "virtual classroom", where interpretive debate is an essential feature. Potter finds that there are distinct advantages in computer-mediated discussion over the conventional face-to-face variety. Instead of the time-honoured classroom discussion where often two or three speakers dominate, with the rest of the class merely listening, Potter observes that computer-mediated discussion encourages much fuller participation, and tends to shift the role of the tutor towards that of an equal participant:

> The decentering of the teacher that occurs in this type of classroom creates a space for students' minds to develop. If you are not providing some form of computer-mediated conversation for your students, if only an out-of-class newsgroup, you are missing one the primary experiences of late 20th-century literature teaching. (Potter 1996: 185).

Collaborative writing may also be enhanced by the unique blend of spoken and written discourse that occurs in computer-mediated discussion where, unlike a conventional discussion, the exchange of ideas is not lost but remains available for consultation and re-editing by all the participants as the process continues. The value of archived discussion is

---

[5]   MUD - "multi-user dimension", a text based game that proceeds according to choices made by the player/s. MOO - "object-oriented MUD" often used for teaching purposes.

supported by the findings of Katchoff and Ryan (1997) and Cronjé (1998). What also arises from this feature is a viable method for collaborative writing projects where a tutor might assign each member of a group a distinct role function - organiser, recorder, stylist, conceptualiser (Potter 1996: 185) - so that each could contribute to the final product in clearly demarcated ways. Potter also applauds the use of simple text analysis programs like TACT where, for example, a student might examine word usage throughout a text or a number of texts. TACT produces word lists that assist in tracing the occurrence of themes and discursive formations but the student must still group the words into semantic fields and subject the results to critical interpretation.

In Chapter 6 we consider the tagging (encoding) of electronic texts that underpins the analytical capacity of TACT and similar software programs, where the adoption of universally compatible encoding schemes has become a major preoccupation in the area of humanities computing. Chapter 7 takes this discussion further by examining TACT and its application in textual analysis in more detail.

Chapter 6 Encoding

Another significant application of electronic texts is in literary research, where the notion of discourse has increasingly gained importance: "'Discourse' has become one those key notions around which all else is constructed" (Bruce 1993: 360). Correspondingly, electronic text provides an ideal analytical tool for the investigation of the explicit and implicit structures of literary texts. As the textual analysis exercises in Chapter 7 will show, the full texts in the SALIT database provide an extensive and versatile resource for the close study of literature, once the user has mastered the software to tag the texts and then use the marked-up texts for analysis. In this chapter I address the nature of text "tagging" or "encoding" and its importance to the literary researcher before going on give some practical examples of electronic text analysis in Chapter 7.

The appeal of using computers to analyse linguistic structures prompted early projects like the Thesaurus Linguae Graecae (TLG) (Hockey 1994) that had successful results but tended to use idiosyncratic markup systems that were not transferable to texts in other fields. TACT[6] is also weak in this respect, because although it has the capacity to incorporate different tag sets and markup, these may still be arbitrarily defined by the user. If SALIT is to realise its full potential as a desktop digital library, teaching and research tool it must be designed to be compatible with similar resources developed elsewhere. Developments in the area of humanities computing have favoured the adoption of a universally applicable tag set that would enable researchers to share electronic texts. The Text Encoding Initiative (TEI) standard promises just such compatibility, and is described in more detail later in this chapter. But first, let us look more closely at the conditions

---

[6] Text Analysis Computing Tools: Text Analysis software application, developed at the University of Toronto and used for the text analysis in Chapter 7.

surrounding the use of electronic text in academic research.

Any study of electronic texts in the humanities must of necessity be inter-disciplinary. The literature reflects this diversity, showing that research into electronic texts in the humanities has emerged from the collaborate efforts of academics in the Humanities (notably in Classical, Medieval and Biblical Studies), Library Science and Information Technology. Interdisciplinary study is also essential to post-colonial studies in South African literature: Van Wyk lists anthropology, history, psychoanalysis, linguistics and philosophy as important in re-defining South African literary history, serving to "free the researcher from 'national meaning'" (Van Wyk 1996a: 36).

The theoretical framework from which electronic textual analysis has developed is that of computational linguistics - the attempt to produce a complete linguistic description of the language in which the text is written. But literary scholars have adapted the analytical tools of computational linguistics for the purposes of literary analysis and interpretation - in short, to find answers based on countable features in texts rather than on intuition or impression. Whereas linguistic researchers strive for scientific accuracy and the recognition of verifiable patterns in the language as a whole, literary researchers make use of statistical analysis as a means to supplement literary critical interpretation and hermeneutics.

Reference was made in the Introduction to the peculiar applicability of computer-based textual analysis in structuralist and post-structuralist literary investigation. Foucault's formulation of an interpretive literary methodology which consists of a

> method of analysis (the quantitative treatment of data, the breaking-down of the material according to a number of assignable features whose correlations are then studied; interpretative decipherment, analysis of frequency and distribution) (Foucault 1972: 10-11)

is eminently achievable through the application of computer-based methods in deconstructing texts. However, before we examine some of the powerful ways in which concordancing and textual analysis can support post-structural literary research, let us question briefly why these methodologies have not more successfully entered mainstream literary research activity.

Bruce (1993) identifies the key weakness in current literary computing as the tendency for researchers to view computer analysis as a "minor adjunct" to the pursuit of conventional literary analysis while ignoring the epistemological status of the electronic text itself. He argues that literary (humanities) computing needs to be developed within the framework of existing models in text and discourse theory "if it is to play a significant role for scholars and teachers" (Bruce 1993: 359). He suggests the use of computerised analysis within the area of textual theory where "text" and "discourse" can be regarded as being signifying systems, where discourse implies a set of anonymous, historically-situated rules determined by the time and space that determine a specific epoch, and where text is a specific articulation of that discourse. Because discourse emerges through text, analysis of the signifiers and signifieds that make up that text is the most promising use to which the processing strength of computers could be applied. The logical and statistical analyses of textual corpora possible with electronic texts lend themselves to this project. Bruce views the role of computerised textual analysis with mixture of enthusiasm and irony:

> We find ourselves in the rather unusual situation of possessing an exciting methodological tool most of whose practitioners rarely articulate a supporting theoretical model adequate to the phenomenon. (Bruce 1993: 360)

The use of computer-based analysis that fits best with the semiotic investigation of texts is its ability to be almost exhaustive in its examination of a given text or corpus. Where scholars in the past have had to work with relatively small corpora, basing their conclusions

on samples selected by intuitive means and through familiarity with known texts, computer analysis makes it possible to discover discursive regularities by thorough inspection of virtually unlimited corpora.

However, the use of electronic texts in the analysis and criticism of modern texts is still relatively undeveloped, with no definitive journal devoted entirely to the subject (although *Computers in the Humanities* and *Literary and Linguistic Computing* both publish literary research). Indeed, the desirability of establishing it as a branch of literary scholarship outside the mainstream is itself a subject of debate (see below), and there are well-founded arguments against its becoming so (Olsen 1993, Wolff 1994). Being able to use electronic texts for textual analysis is a key feature of the design of the SALIT database and a fundamental assumption in this project is the provision of electronic access to primary and secondary material as an appropriate new approach to literary studies. It is also assumed that the ever-increasing familiarity of learners and researchers with computers will salvage what has tended to be a marginal activity carried out by enthusiasts unversed in the issues of contemporary literary theory and bring it into the mainstream.

The reality is that the majority of university literature departments regard the study of electronic texts as belonging exclusively to a branch of computer science. I hope that the wider use of the SALIT Web, and the increasing impact of electronic text centres generally, will begin to dispel this notion. In the 1970s, Nicholas Negroponte (Hall 1996: 2-3) of the MIT Media Lab, predicted that by 2000 the publishing industry would have merged with the television and communications and computing industries. This convergence of all three into an integrated information industry has already started (an example is the ABC/Microsoft Internet broadcasting merger in 1996). Similarly, our transition as a scholarly community from a print-based to an electronic culture is all but complete. No publication reaches a bookshelf today without first having been an electronic text, if not from the author's word-processor, then certainly during the publisher's preparation of the text for print. The fact

that this text exists in the form of electronic bits is not the concern of the literary researcher, but the linguistic and literary applications of digitised text are very significant. The TACT examples in the next chapter reveal how effectively search and retrieval software can be used for content analysis at the word, phrase and concept levels of representation. Making the distinction between simply reading a text, and subjecting it to analysis is fundamental to an understanding of the full potential of electronic texts in the humanities: considerable research efforts are now under way to enable the electronic representation of texts not only to reflect the layout of the printed page, but to delineate the structures of the text (ranging from chapter and scene divisions to syntax) that may be rapidly analysed to test literary critical assumptions.

Olsen provoked a spirited debate in the pages of *Computers in the Humanities* by ascribing the failure of literary computing to exert an influence on "mainstream" scholarship to its "apparent lack of a ... shift in theoretical models and methods to match the new technology" (Olsen 1993: 311). His views are supported by a quantitative analysis of papers published during the 1980s on computer-based literary criticism (Corns 1991). Corns found that the number of such papers had actually decreased over the period, leading him to conclude that "there is no substantial body of achievement in the field of computer-based literary criticism in English studies" (Corns 1991: 127). Both Olsen and Corns, however, believe that literary computing research has been mis-directed until now, and see electronic texts as a solution to the daunting challenge posed by intertextual analysis. The strength of computer based text analysis does not lie in asking fairly traditional questions of traditional texts, but in matching the shift in perspective seen in historical research, away from political and event based history to "the history of social phenomena and the long term, *la longue durée.*" (Olsen 1993: 312). Surprisingly, very few systematic attempts have been made by literary scholars working from within post-structuralist critical positions to make use of the powerful intertextual analytical capabilities of electronic text. Intertextuality and sign theory encourage research design that exploits the strongest points of electronic text retrieval:

rapid analysis of large amounts of data. Such analysis privileges the manipulation of signs over the individual text and author, allowing the critic to view a literary text as "a tissue of quotations drawn from the innumerable centers of culture" (Barthes 1977: 146). Corns foresees large textual databases allowing the machine to "wrap around any text the intertext that surrounded it at its genesis, or at least a subset of that intertext" (Corns 1991: 129).

Already, large textual corpora, such as the ARTFL database (*Trésor de la Langue Française*) (ARTFL 1996) provide an ideal platform from which to conduct such research . ARTFL comprises transcriptions of an extensive selection of French texts for use with a computer, totalling some 150 million words and representing a broad range of written French from novels and poetry to biology and mathematics stretching from the seventeenth to the twentieth centuries. Olsen calls the computer an "ideal semiotic machine" (1993: 313) and cites an investigation into the changing meaning of words that constitute the discourse of gender over long periods of time, where the computer was used to quickly compare sign use in defined blocks of text. The systematic examination of common words such as *femme* and *homme* uncovered unexpected changes in the linguistic encoding of gender relations that would not have been apparent to contemporary French speakers and writers. A similar project by Wolff (1994), using the ARTFL database was mentioned in Chapter 1. Olsen argues that literary computing should move away from analysis in which it has proved a weak substitute for traditional methods - an emphasis on the author's intention, and the stylistic construction of individual texts - towards using electronic texts for the systematic investigation of the development and transformation of "meaning" systems.

> Barthes' declaration of the importance of linguistics in literary criticism and his reduction of the intentions of the author in understanding a text points to an alternative model of literary computing: computers can be used to isolate and examine the symbolic universe in which an author writes (Olsen 1993: 313).

Foucault's suggested methodologies to investigate discursive formations (see Chapter 1)

reinforce Olsen's argument. By re-orientating its underlying theoretical positions towards intextual semiotic analysis in, for example, "...breaking-down ... the material according to a number of assignable features whose correlations are then studied...." (Foucault 1972: 10-11) literary computing could simultaneously provide an important new perspective on the traditional concept of reading texts, and ensure a central role in intertextual research.

Within the immediate boundaries of the SALIT Web, the potential for intertexual analysis is constrained only by the number of texts contained in the virtual library section. As I demonstrate in Chapter 7, a list of deictic markers, once defined, may be used to analyse texts other than Mda's play alone, and when applied to Pringle's texts confirms Serpieri's observation (Serpieri 1978) that drama has a higher "deictic density" (Elam 1980: 140) than narrative or verse. As the collection grows (there were forty-five full texts in the virtual library at the time of writing) so the scope for intertextual investigation will widen. The CSSALL now employs two research assistants to work full time on the scanning and digitising of texts, and the collection will expand further as post-graduate students begin to add texts of their own.

While the opportunity now exists for textual analysis within the present SALIT collection, and will continue to develop within those boundaries, we cannot afford to ignore the importance of ensuring that the SALIT texts are compatible with other digital collections for future collaborative work. Scholarly work in the humanities based on the use of electronic media must allow for diversity and disparities of technical sophistication and access. Particularly applicable to the creation and analysis of literary texts is the demand made on any method of electronic representation to reflect a multiplicity of intellectual perspectives, especially since the humanities also encourage variety and differences in the expression of ideas. This same inherent complexity challenges efforts to standardise terminology or apply quantitative methods of access and analysis. (Hockey 1995, Marchionini 1995). While Olsen's concerns about the apparent resistence in literature

departments to use electronic texts for research cannot be discounted, the SALIT project is founded on the assumption that they will begin to play a larger role in literary research, and we must move ahead to ensure that the resources and techniques are in place to support such methodologies The response to these difficulties does not lie in quixotic tilting at the relevance of electronic resources in literary study, but rather in becoming familiar with and joining the research efforts of literary computing scholars to shape the development of such resources to the requirements of contemporary literary study. Electronic resources are already everywhere in evidence in academia - from word-processors to bibliographic databases - and a thorough understanding of the international effort to have literary computing serve the humanities in the most comprehensive manner possible, with full interchangeability of data, will enable its adoption as valuable research and educational methodology in South Africa as elsewhere.

Susan Hockey, previous Director of the Oxford Centre for Textual Studies, played an important role in the 1976 Oxford Text Archive (OTA) initiative to "prevent texts from becoming 'lost' once their compilers have finished with them" (Hockey 1996: 55-67). The only attempt to catalogue existing electronic texts using standard bibliographic procedures is *The Rutgers Inventory of Machine-Readable Texts in the Humanities* (Hoogcarspel 1994). Since the end of the 1940s various literary computing applications in concordances, text retrieval, stylistic studies, scholarly editing and metrical analyses produced many of the electronic texts that exist today. Until recently, these were generally discounted as by-products of such activities but now they have taken on new value as rich resources for secondary research.

Humanities source material is highly complex, and digital editions demand that characteristics such as multilingual text, variant spelling, parallel referencing schemes, variant readings and marginalia be accurately represented. Not only is the source material intricate in nature, but the demands made by scholars on the texts is equally variable,

including study for literary, linguistic and historical purposes. One significant initiative to solve these problems is the Getty Information Institute's *Research Agenda for Humanities Computing* project, led by David Bearman. Amongst their criteria for the digital representation of texts were *longevity* (ensuring life for the text beyond idiosyncratic or proprietary software markup) and *interoperability* (compatibility with varying types of software). Similar criteria have been adopted by the UK Arts and Humanities Data Services (AHDS):

> The successful management and integration of a highly distributed collection relies heavily upon our adoption and declaration of certain standards. Technical standards used to store information in machine-readable form (e.g. GIF and TIFF for image files, ASCII for text) and including the formal languages which are used to represent the syntactic and semantic features of a digital resource (e.g. SGML for electronic texts, delimited ASCII files for alphanumeric databases) determine how resources may be migrated through changing technological regimes and delivered to users who will want access to them from a variety of different hardware and software platforms (Greenstein 1996a: 6-7).

Any new undertaking in the area of electronic texts, including the SALIT Web project, must take account of these criteria. Furthermore, the emergence of new electronic texts has also required that commonly-used procedures be implemented for the documenting of these texts. Although the need to separate data from software is now better understood, the current proliferation of CD-ROM "viewing" software packages is posing significant problems of support for libraries as well as delivering only partial solutions to many scholarly requirements. Attention is now turning to research towards more advanced network-based delivery mechanisms.

What are other academic disciplines doing to ensure interchangeability of data? Standardising data for research has been a key development in the natural and social sciences in recent years. For example, a 1992 study by The Association of Research Libraries (Cummings et al. 1992) refers to the current scholarly publishing practice of

GenBank, the national repository for nucleic acid sequence data. GenBank requires the full range of substantiating data be made available electronically at the same time that a paper presenting conclusions appears in print. The study goes on:

> Such a model might apply to humanistic scholarship and scholarly communication as well. Many humanists would be as interested as scientists in having ready access to the full range of primary material underlying scholarly arguments (in this instance, of course, the material is ordinarily different in kind, usually text rather than data). In historical disciplines, for example, one can distinguish between the sources - contemporary chronicles and narrative accounts, letters, diaries, works of art, literature, and music, debit-credit registers, data on demographic trends, government statutes, and so on - and analyses that make use of such material and attempt to package and interpret it in particular ways. It is in the comprehensive assembling of the primary material that electronic information technologies are especially flexible and powerful tools, ... because the new technologies permit one to search the assembled primary material with ease and reorder and reassemble it in ways appropriate to one's purposes (Cummings et al. 1992: 2).

What prevents literary scholars from doing this right now? The fact is, that unlike the geneticists, literary scholars have as yet no fully developed method of interchange: there is no electronic *lingua franca* for textual data as there is for numerical data. The texts contained in ARTFL, for instance, are not immediately comparable with texts contained in other databases (they *can* be compared, but only after extensive manipulation of the encoding). A solution to this problem is the creation of "metadata" or knowledge about the knowledge that can make implicit information more explicit, and hence amenable to interchange, comparison and analysis. Hockey (1995) points to the success of bibliographic and cataloguing data in information science in developing metadata to describe material which is mostly not in electronic form at present. She goes on to describe the social science data archiving community's creation in the 1970s of a system for describing its own datasets. These were originally produced in printed form, and some are now being changed into electronic data. She concludes:

Metadata for electronic textual material is in a much more rudimentary form at present and very few electronic texts have what would now be considered adequate information associated with them. Our understanding of the metadata requirements for images, sound and video lags even further behind (Hockey 1995: 2).

If the potential of electronic texts for literary research is its capacity to analyse large textual databases for the purposes of intertextual investigation, then a priority for research in this area must be the creation of such databases, designed in a way that will represent the text faithfully while conforming to internationally standardised encoding practices. By conforming to such a standard, the SALIT library collection is immediately opened up by becoming part of a much larger literary corpus. To this end I have entered into discussions with the South African Data Exchange (SADA) to include literary texts as part of their network of data archives (Stewart 1998: 89-90). SADA's mission is to promote and facilitate the sharing of data among researchers and between the orginators of data and secondary analysts, and while its resources now consist entirely of social statistics, its UK equivalent The Arts and Humanities Data Service (AHDS - see page 110 above) already aims at preserving electronic resources which result from research and teaching in the humanities. The AHDS identifies and promotes shared standards in order to ensure the widespread exchange of digital information. The Oxford Text Archive (OTA) provides a sliding scale of preferred formats for depositing texts with "plain text" formats at the top of the list and proprietary binary files (such as WordPerfect and MS Word and Adobe Acrobat formats) at the bottom because of the lack of certainty about the long-term future of the readers required to display them. The OTA's preferred list is as follows:

SGML TEI or TEI Lite encoded texts are considered most suitable. Also acceptable, in order of preference are: Data encoded using other SGML DTDs (including HTML or XML), LaTeX or TeX, Rich Text Format (RTF), and finally ASCII or ("plain text").

My analyses using TACT (see Chapter 7) were performed on ASCII text, with COCOA[7] markup (a predecessor of SGML/TEI). For the SALIT collection to achieve true exchangeability they require encoding in TEI format.

I will now go on to give a short account of the Text Encoding Initiative (TEI) and consider its applicability to the SALIT project firstly by providing a background to the development of the TEI, and then going on to describe the principles underlying the encoding scheme. The project originated at a planning conference convened by the United States Association for Computers and the Humanities (ACH) at Vassar College in 1987. The consensus of the Vassar conference participants, representing expertise from universities, learned societies, and text archives from North America, Europe, Israel and Japan, was that a common framework for a common set of guidelines for the encoding of machine-readable text was necessary and feasible. Amongst the key factors that contributed to this consensus was the recently developed Standard Generalised Markup Language (SGML), defined by the international standard ISO 8879, that appeared to provide an invaluable tool for developing a simple, flexible, extensible encoding scheme capable of satisfying the widely varying needs of textual researchers.

SGML is not in itself an encoding scheme but a framework within which encoding schemes (tag sets) may be developed. Because multiple tag sets may be used in the same texts, SGML-based encoding schemes can easily support a diversity of opinions about the basic text features to be tagged. SGML is device-independent and is supported by software products from an increasing number of vendors. (The Folio Infobase software, that we have chosen for the *South African Literary Encyclopaedia CD-ROM* project supports SGML, as do recent releases of popular word-processing packages such as WordPerfect and MS

---

[7]   Pre-SGML tagging developed for the Oxford University Press Micro-OCP concordance program; also used in TACT.

Word). SGML can handle any natural language for which an electronic representation exists and is by design application-independent. This means that it enables the same text to be used for many different applications in different environments.

The publication in 1994 of the *Guidelines for Electronic Encoding and Interchange* (TEI), as the culmination of the first six-year phase of the project provided an appropriate way to consistently and comprehensively encode a wide range of text types and textual features. (Ide and Véronis 1995a). In other words, the TEI encoding of text separates the content (the words on the page once the text is printed) and markup (information about the structure of the text, its division into chapters, lines or scenes). The "reveal codes" feature of a typical word-processor will show similar markup of the text. However, SGML (Standard Generalised Markup Language on which the TEI is based) differs significantly from that used by word-processors in that the markup is defined in abstract terms, rather than direct processing terms. Apart from its appropriateness in the rendering of literary texts, SGML/TEI encoded text is software-independent, ensuring its portability from one word-processor (or any text viewing platform, including Internet browsers) to the other. Liberation from the constraints of proprietary software is crucial to the use of electronic text for academic purposes; Sperberg-McQueen, co-leader of the TEI project quotes the slogan: "With SGML, you own your documents; without SGML, your documents are owned by people in Orem, Utah, or in Redmond, Washington. Which do you prefer?" (Sperberg-McQueen 1995: 3). His references are to the two most widely used word-processors, WordPerfect and Microsoft Word. (The software industry is rapidly adjusting to its perceived market, however. Since 1995, both word processors have incorporated SGML/TEI authoring and editing features.)

Application independence is an important consideration for the SALIT Web project, as it ensures that the content of the publication will outlast the particular software chosen to display it on the CD-ROM. Although the full texts in the SALIT virtual library are currently

either in ASCII[8], MS Word or Folio Views formats, all are readily compliant to TEI markup planned in the next phase of the project. Another crucial advantage of this policy is the flexibility it allows in giving users access to the text for their own research and study purposes.

The *Guidelines for Electronic Encoding and Interchange* (Sperberg-McQueen and Burnard 1994) represent the output of a number of specialist work groups. The work groups made recommendations in their areas, either directly where the area was well-defined, or indirectly by sketching out a problem domain and proposing other work groups which needed to be set up within it. The groups included: character sets, text criticism, hypertext and hypermedia, mathematical formulae and tables, language corpora, physical description of manuscripts, analytic bibliography, general linguistics, spoken texts, literary studies (which subsequently expanded into groups for drama, verse and literary prose), historical studies, machine-readable dictionaries and computational lexica. Scholarly discussions of encoding debates within many of the disciplines represented in the working groups can be found in a special edition of *Computers and the Humanities* (Ide and Véronis 1995b) devoted to the background and contexts of the Text Encoding Initiative.

As regards to the concepts of markup (text encoding) and SGML, Sperberg-McQueen (1995: 3) draws a primary distinction between SGML markup, which is abstract or content-oriented, and proprietary markup such as that used in the word-processing programs mentioned above, which is procedural in nature. Whereas SGML will mark a phase as "technical" and leave the formatting of the phrase to a standard stylesheet, a word-processor will typically mark the phrase "italics" regardless of content. Both will produce

---

[8]   ASCII (American Standard Code for Information Interchange). A standard code for representing characters as numbers that is used on most microcomputers, computer terminals, and printers. (See Glossary - Appendix F.)

the same printed (or displayed) final output, but the SGML version remains device-independent and allows not only for variable formatting (bold capitals, coloured letters, or underlining rather than italics) but also for context-sensitive analysis of the text (as, for example, a count of *technical terms* rather than *titles* that occur in a given text).

The TEI tagset is defined in terms of SGML, and all TEI-conformant documents must also conform to SGML, typically accompanied by system-independent documentation of the markup. The three basic forms of markup within SGML were outlined: *declarations, entity references* and *tags*. If the TEI tag set is used, then all *declarations* are included, with the possible exception of additional comments, but for most documents *entities* (named portions of documents) would need to be defined. An example of such an entity would be the inclusion of non-standard characters in the text: the entity *eacute* means the character "e with an acute accent" (é). *Tags* are used to mark the beginning and ending parts of a document; these parts are known as *elements*. Tags are indicated by the use of angle brackets, the end-tag having a slash after the opening angle bracket. In the following example, the sentence is marked as a quotation by the start-tag and end-tag which surround it; *quote* is an element type defined by the TEI.

<quote>L'état, c'est moi.</quote>

One can recognise the similarity between these tags and the COCOA tags used for the markup of Mda's play *The Hill* illustrated in Chapter 7. All SGML documents must conform to a Document Type Definition (DTD) which includes declarations of all the entities, elements and attributes allowable in the document. The TEI DTD, developed by the wide range of discipline specialists described above, is uniquely suited to the requirements of humanities texts. It is helpful to use an analogy here: the TEI DTD, and others such as HTML3 (used for internet documents) are to SGML as individual stylesheets would be to the Modern Languages Association (MLA) stylesheet (including, as it does, a set of

conventions for the presentation of research).

During the course of this project, I have observed the increasingly widespread adoption of TEI markup for electronic texts in the humanities. The TEI DTD (and TEI Lite, a shorter version) is now used by most of the major providers of electronic texts, including those provided by publically accessible collections such as the University of Virginia, scholarly projects like the Brown Women Writers Project, and commercial publications including the Chadwyck-Healey Full Text databases. As mentioned earlier, data exchange services such as the Arts and Humanities Data Service in the UK also prefer TEI markup. The benefits to the user of the SALIT Web, once the full texts are marked up in accordance to the TEI Guidelines are clear.

Balancing the demands of marking up texts in TEI against the equally pressing demands of designing and building the database structure of SALIT as a whole was a major challenge during this project. Working within the constraints of limited time and assistance, my compromise solution, in order to deliver both the database and make a start on building the full text collection, was to train a research assistant in the rudiments of text scanning and the preparation of the basic ASCII files, allowing her to use her existing editing and proof-reading skills to ensure the integrity of the digital texts. The next step, training in TEI markup, was postponed to the next phase of the project as it was also necessary to develop simultaneously the bibliographic "backbone" of the database which grew from 4000 to 35000 title entries over a three-year period, each one having to be individually entered. In addition, the image library (consisting at present over 500 images, mostly photographs of the authors) had also to be entered into the database. An account of this process is covered in greater detail in Chapter 10, but the point is that the TEI markup had to be sacrificed temporarily in order to ensure the completion of the overall web in which they would be housed.

Nevertheless, the relationship between various aspects of the SALIT project and the TEI was closely investigated, as my discussion of this issue in Chapter 10 will show. As far as the individual full texts are concerned, experiments with TEI markup are interesting to compare with the TACT-based markup used for the Mda and Pringle texts.

```
<author Thomas Pringle><work African Sketches: Part 1: Poems Illustrative of South
Africa>
<edition William Hay, 1912><<scanned 1996/7 by Rashmi Jadhundan; proofread and
edited by Rashmi Jadhundan and Graham Stewart. CSSALL>>

<poem 22>< THE FORESTER OF THE NEUTRAL GROUND. A SOUTH-AFRICAN
BORDER-BALLAD.>

<L 1.> We met in the midst of the Neutral Ground,
<L 2.> 'Mong the hills where the buffalo's haunts are found;
<L 3.> And we joined in the chase of the noble game,
<L 4.> Nor asked each other of nation or name.
```

Figure 29: TACT markup: Pringle

```
<!ENTITY author "Thomas Pringle"
<! – reproduced from van Wyk, J, Conradie P, Constandaras N. (1988) SA in Pöesie/
SA in Poetry. Pinetown: Owen Burgess. pp. 33-35. From Pringle, T. Poems Illustrative
of South Africa. Cape Town: C. Struik ->


<!ELEMENT anthology - - (poem+)>
<!ELEMENT poem - - (title?, stanza+)>
<!ELEMENT title - 0 (#PCDATA)>
<!ELEMENT stanza - (line+)>
<!ELEMENT line 0 0 (#PCDATA)>


<anthology>
<poem><title>The Forester of the neutral ground - A South African border-
ballad</title>
        <stanza>
<line> We met in the midst of the Neutral Ground, </line>
<line> 'Mong the hills where the buffalo's haunts are found; </line>
<line> And we joined in the chase of the noble game, </line>
<line> Nor asked each other of nation or name. </line>
</stanza>
</poem>
</anthology>
```

Figure 30: TEI markup: Pringle

The overall similarities in the use of angle brackets, for instance, is evidence of features of
the COCOA markup that have been inherited by the newer TEI scheme. But the "nested"
elements (anthology, stanza, poem, line) and their formal declaration in the DTD (the
<!ELEMENT statements), required by the TEI encoding, hint at the much more robust
structural coding of the TEI fragment. It is this conformity with the TEI standard that

ensures the exchangeability of the text, and its compatibility with texts in any of the large electronic text collections referred to earlier. Once the SALIT texts are encoded according to the TEI standard, TACT may still be used to perform the types of textual analysis illustrated in Chapter 7 as TACT now includes an SGML import option that accommodates TEI markup. As the TEI encoded electronic texts have become more widely available, programs have appeared that will for example, convert TEI files into HTML for Internet distribution or into word-processing applications like WordPerfect.

A similar comparison of TACT vs TEI markup may be seen in extracts from Mda's *The Hill*, below:

Figure 31: TACT markup: Mda's *The Hill*

```
<author Zakes Mda><work The Hill><edition Unknown>
<<scanned 08/98; preliminary edit by Graham Stewart 08/98>>
<<proofread & corrected by Graham Stewart 09/98>>
<<line-numbers added by G.S. 09/98>>
<<TACT label and speaker abbreviations added, and >>
<<further corrections to the text G.S. 09/98>>
<scene -><L ->«p»
<tt frontmatter:title>
ZAKES MDA
THE HILL
<tt frontmatter:characters>
CHARACTERS
<<abbreviations and alternate forms given in braces following>>
THE NUN {NUN}
THE MAN {MAN}
THE YOUNG MAN {YMAN}
THE VETERAN {VET}
```

FIRST WOMAN {1WOM}

SECOND WOMAN {2WOM}

THIRD WOMAN {3WOM}

<tt play><scene 1>

<set The Hill>

<tt stagedir> {A silhouette of a NUN in full habit. She is sitting on a rock and seems to be in meditation. The general atmosphere of the place is that of a graveyard and the rock she is sitting on gives one the idea of a headstone. There are rocks all around her and these enhance the romantic feeling of a graveyard. Even in the dimness one can see that there is a rosary dangling from her clasped hands and she is also holding a big flower -- most likely a plastic rose.} <tt play>

«NUN» <tt stagedir> {reciting in monotone} <tt play> Mea culpa, mea culpa, mea maxima culpa. <tt stagedir>{Unearthly voices, also in monotone, slowly repeat every syllable after her.}<tt play>

«UNEARTHLY VOICES» Mea culpa, mea culpa, mea maxima culpa. Mea culpa, mea culpa, mea maxima culpa.

<tt stagedir> {A MAN enters. He has a blanket over his shoulders. He is panting heavily because he has been running. He perches himself on a rock not far from the NUN, giving her his back. NUN continues with her 'mea culpa' but now in lower monotone.} <tt play>

«MAN» Mother, cleanse me! Wash my sins.

Figure 32: TEI markup: Mda's *The Hill*

<div0 type="scene" n= "1">

<stage> A silhouette of a NUN in full habit. She is sitting on a rock and seems to be in meditation. The general atmosphere of the place is that of a graveyard and the rock she is sitting on gives one the idea of a headstone. There are rocks all around her and these enhance the romantic feeling of a graveyard. Even in the dimness one can see that there is a rosary dangling from her clasped hands and she is also holding a big flower -- most likely a plastic rose. </stage>

<sp>

<speaker>NUN</speaker>

<stage> reciting in monotone </stage>

<l n=1>Mea culpa, mea culpa, mea maxima culpa.</sp>

<stage> Unearthly voices, also in monotone, slowly repeat every syllable after her.</stage>

<sp>

<speaker>UNEARTHLY VOICES</speaker>

<l n=2>Mea culpa, mea culpa, mea maxima culpa. Mea culpa, mea culpa, mea maxima culpa.</sp>

<stage> A MAN enters. He has a blanket over his shoulders. He is panting heavily because he has been running. He perches himself on a rock not far from the NUN, giving her his back. NUN continues with her 'mea culpa' but now in lower monotone.</stage>

<sp>

<speaker>MAN</speaker>

<l n=3>Mother, cleanse me! Wash my sins. </sp>

As with the Pringle example in Figure 30, the more tightly specified nested elements of the TEI encoding can be seen. The flexibility of the TEI markup would allow for the tagging of deictic markers (see Chapter 7), nested within the existing encoding for the drama structure. It must be emphasised here that none of this encoding is meant to appear in a display of the text intended for normal *reading*. The value, as has been demonstrated, lies in the user's being able to locate explicit structural items for analysis. When the TEI text is viewed on a web browser, or within a word processor using an SGML conversion facility, the tags vanish, leaving the plain text alone for easy reading.

In the next chapter I go on to trace the development of text tagging from its origins in the making of concordances to the way it is currently employed in literary analysis. Practical applications using TACT illustrate the potential of a text analysis program as a methodological tool for the literary researcher.

Chapter 7 Concordancing and Textual Analysis

To what extent can a textual analysis program add value to literary research? The practical applications of TACT described in this chapter are an attempt to explore ways in which appropriately tagged electronic texts may contribute to analysis and interpretation in South African literature study. Examinations of electronic texts by Pringle (1834) and Mda (1995) drawn from the SALIT Web library, show that computer-based analysis can provide a "key empirical and heuristic tool" (Bruce 1993: 361), but pose difficulties for researchers with only rudimentary computer skills. The exercise also demonstrates that the use of a program like TACT must be seen only as a means to an end: these analyses all require a certain amount of manual intervention.[9]

An ideal project to test the capacity of computer-based textual analysis in post-modern literary study would be one similar to that carried out by Marc Angenot in his *1889. Un état du discours social* (1989). Angenot's approach, transposed into a South African literary context suggests a synchronic discursive history in which the interactions of various discourses could be examined within a specific temporal limit (e.g. 1820 to 1830) by searching as comprehensive a set of texts as possible for the determination of thematic, rhetorical, lexical, structural and discursive regularities. As Bruce has argued (see above), computer-assisted analysis "should be a key empirical and heuristic tool" for such a purpose. Electronic text analysis would also ensure that large numbers of texts could be analysed with great precision, reducing the likelihood of false conclusions drawn from unrepresentative samples of data.

---

[9]   See McCarty's caveat on page 146.

Although a project of this magnitude lies beyond the scope of the present study, some indication of the direction such research could take can be found in the word-frequency analyses of Thomas Pringle's *African Sketches* below. Discursive regularities in contemporary South African travel writing by other authors (Burchell, and the Rev. Campbell, for instance) could be revealed by applying similar statistical methods to a corpus expanded to include their works in addition to those of Pringle, and, given the availability of electronic text versions, other writers of the era.

The first application of text analysis in literary computing was in the area of word indexes and concordances. Concordances in textual study have a long history, traceable back to Conrad of Halberstadt's 1286 concordance to the Vulgate Bible devised as a biblical teaching tool. The first modern concordance is attributed to Alexander Cruden, whose concordance to the Bible appeared in 1736. Computing humanists today use concordancing for essentially the same purposes as the 13th Century Dominicans: "to discover patterns of coherence in a text or textual corpus" (McCarty 1996: 2) The first use of computer technology for concordancing was Father Roberto Busa's investigation of the philosophical and theological idea of inwardness in the writings of St. Thomas Aquinas, a project that began in 1949 and led to the publication of the sonorously titled *Thomae Aquinatis Opera Omnia cum Hypertextibus* on CD-ROM in 1992 (Busa 1992).

A concordance is essentially a comprehensive index of the words used in a text or a corpus of texts, but while a word index lists all the words in a text, giving the locations of the occurrences of each word, a concordance is a far richer resource, including some context for the word, for example, ten words on either side. Often a word-frequency count is also given, listing the number of times each word occurs in the text (as in the word frequency list from Pringle's work - see page 132 below). Therefore the concordance offers the literary researcher considerably more than a simple index: when you look up a word in a concordance you are able to examine a gathering-together of *all* the usages of the word, as

well as being able to compare the different contexts within which the word is used. By isolating the word and its contexts, you can more readily discover significance in the patterns in which the word has been used, and in the characteristic language of the text. While the concordance can show how often a particular word is used, it can also reveal which words are *not* used, which can contribute just as valuably to the interpretation of the text.

A short digression on the physical characteristics of an electronic text is necessary before we proceed. It is important to note that nearly all work in this area has used electronic text *transcriptions*, rather than digitised *images* of the text: a transcription allows analysis and manipulation of the text (counting, locating references, etc.), whereas a digital image is analagous to a photographed snapshot of the page. Text preserved in a digital image may be read like a conventional printed page, but is not readily accessible to computerised textual analysis. (There *are* systems for subjecting digital images of text to search and retrieval, but these are relatively undeveloped at present. An example is the searchable electronic *Palestine Post* Humanities Computing Project at Tel Aviv University, see Zweig 1997).

The following discussion, then, assumes that the researcher has access to an electronic text transcription. *The Oxford English Dictionary* definition of a concordance is "An alphabetical arrangement of the principal words contained in a book, with citations of the passages in which they occur" (OED 1994 s.v. 6.b.). But McCarty (1996: 1) amplifies this definition by specifying four major types of concordance: the verbal, contextual, glossarial and conceptual. The verbal, sometimes known as the *index verborum*, consists of a list of word-forms with citations to the text, but omitting context. As its name implies, the contextual concordance includes fragments of actual text (as in the KWIC - keyword-in-context output shown below from Pringle's *Poems Illustrative of South Africa* - see page 137). The "contextual" concordance corresponds most closely to the OED definition, while the "glossarial" form has a lexical structure, with dictionary-like headwords, arranged by

grammatical or orthographic lemmas (or "word stems"). Although all of these can, and have been used for literary analysis, it is the last type, the conceptual, that offers the most scope for computer-based literary analysis.

McCarty's last type, the "conceptual" concordance is similar to a thesaurus or "concept dictionary", where the textual word-forms may not appear in the quoted text fragments, but where there is a conceptual connection to the listed headwords. However, the conceptual concordance is also the most controversial of the four types, especially if produced in printed form. A published concordance will inevitably reflect the compiler's choice of lexical items, and the interpretive bias of the compiler will be particularly limiting where the grouping of words follows the compiler's own reading of their associations, rather than one based purely on word-forms. For a researcher hoping to use such concordances to test his own textual assumptions, fixed listings can pose an almost insurmountable barrier to interpretation. Within the suite of resources that constitute the SALIT project, all of these types are available to the user in some form: the main user interface, for example, is predominantly "conceptual", reflecting decisions made by the editors (e.g. the keyword, "guide" and "tour" options). But the user is also able to search the database without any editorial mediation and may search for specific words or from alphabetically-arranged indices of author names, titles and full texts. Within the full texts themselves there is no authorial intervention at all at present - it is left to the users to utilise the text analysis software in any way they choose.

Although McCarty's four-type taxonomy outlined above is useful in describing the intentions of the concordancer, the development of electronic text analysis tools since the 1980s (e.g. TACT) means that in practice, the first two types of concordance can be generated automatically (in SALIT both within the database and individual full texts), while the last two can be easily derived through time-consuming, but straightforward hypertext markup, thus offering varying levels of entry for the user. At one end of the scale, an

undergraduate user may be guided at first by pathways "hard-wired" into the resource by the editors, while at the other, experienced researchers may search the material free of the pre-determined interpretive schemes that are inescapable in printed concordances.

McCarty (1996: 12) explains that once the process of making a concordance had been automated (by the design of a computerised algorithm), then it became possible to use additional algorithms to analyse textual data in other ways. Textual analysis, then, allows the researcher to model ideas about how texts work.

To see how such a tool could be of use to the literary researcher, it is helpful to consider the questions that TACT (Text Analysis Computing Tools) (Bradley, Presutti and Stairs 1994) can answer. John Bradley of the University of Toronto, principal designer of the progamme, puts it this way:

> As well as allowing you to choose words from the vocabulary as a way of creating your selected list, TACT lets you specify rules for the selection of words or textual positions. There are several different types of selections TACT can do for you:
> 1. you can give it a pattern to use to select word forms;
> 2. you can let it find words that are similar to a given word;
> 3. you can ask it to select positions in the text based on co-occurrence of two events: the last example ... shows a fragment of the result TACT produced when asked to find all places in the text where a 'love' word was close to a 'death' word;
> 4. you can refine your word selection by structural information: "find all 'love' words spoken by Juliet";
> 5. and by frequency: "find all words that occur more than 100 times".
> (Bradley 1991: 3)

Textual analysis of Pringle's *African Sketches* using TACT was carried out using two different approaches. In the first, a frequency-ordered word list was generated, collocations (patterns of words occurring within nine words of each other) were determined, and all similes with the patterns "like ..." and "as ..." were transcribed. The intention was to use the

word frequency patterns to suggest thematic categories, and so uncover dominant concerns in Pringle's work. In the second approach, some of A.E. Voss' conclusions in his paper "'The Slaves Must be Heard:' Thomas Pringle and the Dialogue of South African Servitude" (Voss 1990) are subjected to a limited computer-based analysis (specifically the Pringle texts) to test his assumptions about the significance of images of servitude and slavery in English narrative and discursive texts during the first and second British occupations of the Cape.

The two approaches, one proceeding from an empirical listing of words and phrases to arrive at possible analytic directions (inductive), and the other, to support or refute "intuitive" interpretation based on literary theoretical assumptions with textual references found without the help of a computer (deductive), were used to demonstrate the strengths and the limitations of computer-based textual analysis. The following diagrams summarise the two approaches as "inductive" and "deductive".
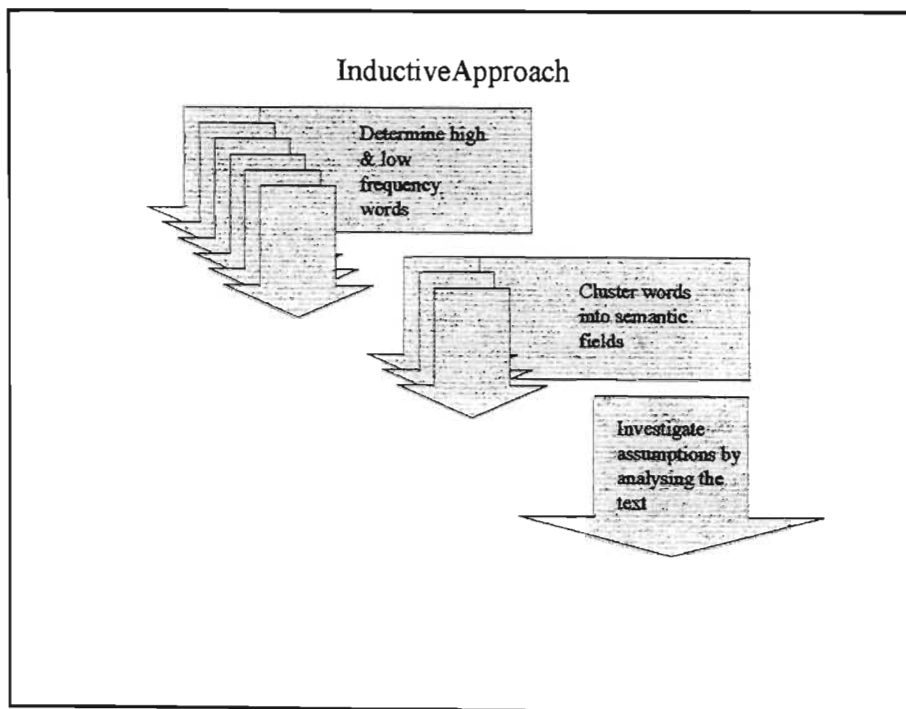
Figure 33: Textual analysis - Inductive Approach

The first approach to the Pringle text was based on a model provided by Mills and Chandramohan who used TACT to teach Conrad's *Heart of Darkness* to BA (Honours) students at the University of Luton. They argue that computerised textual analysis develops students' awareness of both the literary and the linguistic characteristics of a text.

> Computer text analysis and retrieval software enables learner-centred, conjectural learning from an empirical basis ... An empirical basis is essential if a theory is to be credible and the computer provides the means to collect data not only more accurately but also for much longer texts than would prove feasible by human endeavour (Mills and Chandramohan 1996: 166).

The suitability of computer-based textual analysis to longer texts is clear from the analysis of *African Sketches*, where *Poems Illustrative of South Africa* consists of 42 poems with a total of 2463 lines of verse, and *Narrative of a Residence in South Africa* is 350 pages in length. If one discounts the scanning, proofing and encoding of the text (amounting to several weeks of concentrated work), the various forms of analysis available in the TACT program can be performed almost instantaneously. Hence the emphasis in this study on the importance of building as comprehensive corpus as possible of South African literature. As I have attempted in a limited way here with the Pringle texts, Mills and Chandramohan used TACT both to find patterns of stylistic, lexical and grammatical nature, and to reconfirm the conclusions drawn by literary critics by traditional analytic means. To begin with, the students at Luton performed a lexical analysis of *Heart of Darkness*, producing a selected word list of items that occurred between 20 and 99 times. At this point they eliminated frequently occurring "closed class items" (articles, common prepositions, etc.) as well as words of low frequency. From the remaining list of high-frequency items, the instructors guided their students to categorise the words into semantic fields under the headings "Antonyms", "Topography", "People", "Parts of the body", "Activities", "Objects" and

"Miscellaneous". Amongst other things, the students noticed that although the words "man" and "men" appear in the high-frequency list (some 166 times), "women" does not. The authors link this observation to an existing critical commentary on the novel which cites the background role of the women in the story as one of its strengths because in his other novels, "Conrad's greatest fault [...] is his inability to create convincing women characters" (O'Prey 1983: 22).

In my analysis of *African Sketches*, the two texts *Poems Illustrative of South Africa* and *Narrative of a Residence in South Africa* were at first examined separately. This decision was influenced by the differing genres the two works represent, and to make it easier to isolate the different lexical patterns of the two for later comparison. The markup of the texts was performed on the plain text version produced by scanning the print versions (for a fuller description of this process, see Chapter 4). The markup in the *Poems Illustrative of South Africa* consisted of TACT tags indicating "$poem" (the reference number assigned to each poem, e.g. "Afar in the desert" is no. 2) and "$lineno" identifying each line of verse within the poem. This markup ensured that all later analyses of individual words, words in proximity (KWIC - key word in context) and collocations would contain direct references both to the original and to the Folio Views version (the hypertext edition produced for inclusion in the SALIT database). *Narrative of a Residence in South Africa* had tags for $chapter and $pageno (the page numbers corresponding to those in Lewin Robinson's 1966 edition). Both texts were then converted to textual databases with the TACT Makebase program.

The first set of analyses was a word frequency count using TACT's TACTfreq program that produced a list of all words in each of the databases: poemsill.tdb (*Poems Illustrative of South Africa*) and narrativ.tdb (*Narrative of a Residence in South Africa*). Both lists were saved as *.alp files that could be opened as text files in MS Word. Tables of the most commonly occurring words of significance were then prepared as follows:

1. In the Word document, the process of excluding "closed class" and low frequency items began with the marking of each selected "open class" word with an @ symbol.

2. The second step is using the Word find/replace function to mark each @ entry as a Header 1 using the format style option.

3. Using insert index/tables choose TOC (table of contents) with "show page numbers" unchecked and "show levels" = 1.

4. The newly generated TOC is retained while the rest of the word list is deleted.

5. The process is completed by changing the style to "normal" and deleting all the @ marks. The file is then saved with a new name, retaining the original *.alp file intact for further analysis.

The following tables show the results of this process:

Poems Illustrative of South Africa **Selected word frequency list**			
40 wild	12 flock	9 afric's	2 natives
31 heart	12 found	9 child	2 slave
30 round	12 friend	9 dread	2 slavery
28 high	12 friends	9 fell	1 amakosa's
25 far	12 god	9 grey	1 black-skinned
24 still	12 herds	9 hills	1 bondmen
23 desert	12 life	9 lone	1 british
23 men	12 long	9 scottish	1 bushmen
21 hand	12 race	9 silent	1 bushmen's
21 land	12 river	9 song	1 cafferland
19 old	12 soul	9 vale	1 caffers
17 away	12 sound	9 wood	1 christ
17 fear	11 bright	9 woods	1 enslaved
17 green	11 fair	8 above	1 fiends
17 light	11 flocks	8 bold	1 gunja
17 man	11 hear	8 bones	1 insane
16 dark	11 past	8 dear	1 laboured
16 good	11 prey	8 look	1 laughs
16 home	11 pride	8 lost	1 lonely
16 mountains	11 rocks	8 man's	1 nature
15 afar	11 waste	8 other	1 negro
15 love	10 blue	8 stream	1 poetic
14 brown	10 christian	7 native	1 romance
14 call	10 deep	5 black	1 romantic
14 proud	10 den	5 bush-boy	1 rousseau
14 see	10 heard	5 bushman	1 servitude
14 well	10 only	4 serve	1 slaves
14 white	10 sky	4 servile	1 slavish
13 alone	10 sons	3 caffer	1 solitary
13 eye	10 time	3 hottentot	1 somerset
13 savage	10 voice	2 bondman's	1 sublime
12 blood	10 wake	2 boor's	1 tribe
12 day	10 young	2 bushman's	

Narrative of a Residence in South Africa **Selected word frequency list**			
250 colony	83 chiefs	54 either	1 lone
248 government	83 land	54 home	1 lonely
214 river	82 captain	53 circumstances	1 lonesome
177 colonial	80 order	53 friend	1 luck
156 time	80 period	53 others	1 mad
151 frontier	80 wild	52 life	1 madly
142 party	79 might	52 policy	1 madness
140 colonists	77 british	51 commissioners	1 philosophy
135 caffers	75 native	51 occasion	1 poem
132 hottentots	72 settlers	51 public	1 poet
131 town	71 african	51 together	1 poetic
129 boors	70 england	50 must	1 poets
128 hottentot	70 mountains	50 nearly	1 poet's
123 people	70 territory	50 white	1 profi
117 cattle	70 valley	49 friends	1 repress
111 years	69 character	1 amuse	1 repressed
109 chief	69 day	1 amused	1 repression
107 caffer	69 english	1 amusements	1 reputation
107 far	69 new	1 bantu	1 servility
105 state	67 missionary	1 christ	1 signifies
102 man	67 natives	1 democratic	1 stubborn
102 place	65 service	1 emotion	1 stunted
99 governor	64 stockenstrom	1 emotions	1 tolerated
93 somerset	63 family	1 happy	1 trouble
91 should	63 race	1 hobhouse	1 ugly
87 against	62 dutch	1 immigrants	1 unfeeling
87 following	62 inhabitants	1 indentured	1 uninhabited
87 good	60 children	1 insane	1 unnatural
86 settlement	59 tribes	1 insanity	1 unsophisticated
85 present	57 account	1 joyful	1 wealthy
85 see	56 night	1 laugh	1 write
85 south	55 condition	1 laughable	1 zoola
84 africa	54 driven	1 laughed	

Semantic fields for each text were identified as follows:

Poems Illustrative of South Africa	Narrative of a Residence in South Africa
Antonyms light/dark, old/young, fear-dread/bold	**Antonyms** day/night, wild/order, should-must/might
Topography desert, (wild), land, home, mountains, river, rocks, sky, hills, vale, wood(s)	**Topography** river, frontier, place, settlement, south, africa, land, england, mountains, territory, valley, home
People men & man, (savage), friend(s), god, race, sons, child	**People** party, people, stockenstrom, family, inhabitants, children, friend(s), others, together, somerset, man
Activities call, see, found, hear(d), silent	**Activities** see, driven
Objects flock(s), herds, sound, prey, den, wake, song	**Objects** cattle
Ethnic groups	**Ethnic groups** colonists, caffers, hottentot, boors, british, native(s), settlers, african, english, tribes, dutch
Body parts heart, hand, eye, blood, soul, voice	**Political/state** chief, governor, captain, missionary, commissioners
Emotions fear, love, proud, alone/only/lone, soul, pride, dread	
Colours green, brown, white, fair, blue, grey	
Miscellaneous savage, wild, round, high, far, afar, still, away, light, bright, time, day, life, long, past, waste, deep, only, fell, christian	**Miscellaneous** time, years, present, period, character, against, good, public, new, service, account, condition, circumstances, policy, decision, nearly, life

Although the scope of this chapter precludes further analysis of the data, it is worth noting that the word lists suggest several interesting lines of investigation - for instance the prevalence of the imperative and directive tone evident in the *Narrative* revealed by the high frequency of "should, must" and "might"; and the relative infrequency of ethnic labelling and words expressing emotion in the *Poems* as compared to the *Narrative*.

The second approach which I have called "deductive", proceeds from conclusions drawn by Voss (Voss 1990: 75) that in the work of Pringle and some of his contemporaries (a) slavery is shown to debase both the exploited and the exploiter, (b) "the slave never speaks" - servitude is rendered in the third person, and (c) there are unresolved ambiguities in the concepts of "servitude" and "slavery". Voss' insights may be tested by using the text database to query the use of words related to slavery.

The method of analysis may be represented diagrammatically as follows:
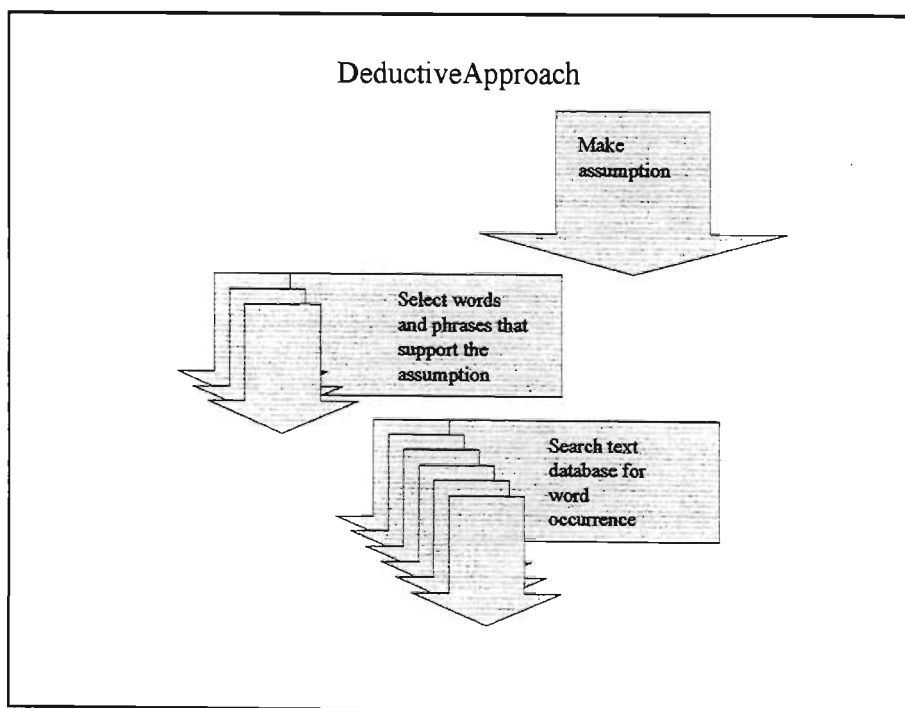
Figure 34: Textual analysis - Deductive Approach

Words and phrases in African Sketches that relate to the idea of slavery and servitude were identified, and a query group created in the text database to find all instances of these words. The query file consisted of the following words:

Query file "poemsill.que": slave: bond.*,serv.*,labour.*,thral.*

The query returned the following list (numbers in brackets indicate the frequency of occurrence in the text):

bond (1)	free (7)	freeman's (1)	servest (1)	slave (2)	thraldom (2)
bondmaid (1)	freeborn (1)	laboured (1)	service (1)	slavery (2)	thrall (2)
bondman's (2)	free-born (1)	serve (4)	servile (4)	slaves (1)	
bondmen (1)	freedom (4)	serves (1)	servitude (1)	slavish (1)	

The relevance of each of these words to the assumptions being tested could be considered by using the display options available in TACT, i.e. "KWIC" (keyword-in-context), variable display and the full text, but for the purpose of demonstration, the following report was generated, based on the word-search (only three of the forty-two entries are shown here - the references on the right are to the poem and line numbers Poem 38 is "The Cape of Storms", Poem 26 is "The Hottentot" and Poem 29 is "Slavery"):

<L 4.> To stanchest mariner and stoutest bark;

<L 5.> And though along thy coasts with grief I mark

<L 6.> The servile and the slave, and him who wails

<L 7.> An exile's lot-and blush to hear thy tales

<L 8.> Of sin and sorrow and oppression stark:-

———————————————————————————————————— 38.6.

servitude (1)

<L 8.> Submissively his freedom and his lands.

<L 9.> Has he no courage? Once he had-but, lo!

<L 10.> Harsh Servitude hath worn him to the bone.

<L 11.> No enterprise? Alas! the brand, the blow,

<L 12.> Have humbled him to dust-even hope is gone!

———————————————————————————————————— 26.10.

slave (2)

<L 3.> Which taints with leprosy the White Man's soul,

<L 4.> Not less than his by whom its dregs are quaffed.

<L 5.> The Slave sinks down, o'ercome by cruel craft,

<L 6.> Like beast of burthen on the earth to roll.

<L 7.> The Master, though in luxury's lap he loll,

———————————————————————————————————— 29.5.

Figure 35: KWIC display of "slavery" query (poem and line number on the right)

TACT provides a distribution display that may be configured to provide a graphical display of word occurrence, synchonised to the list of words being queried and so signposting the poems where the word (or concept) is most often referred to :

```
May 16, 1998.

1       |  3
12      |  1
13      |  1
18      |  2
2       |  1
20      |  1
22      |  4
26      |  1
29      |  2
33      |  2
35      |  1
38      |  2
39      |  1
6       |  4
9       |  2

Total: 28.
```
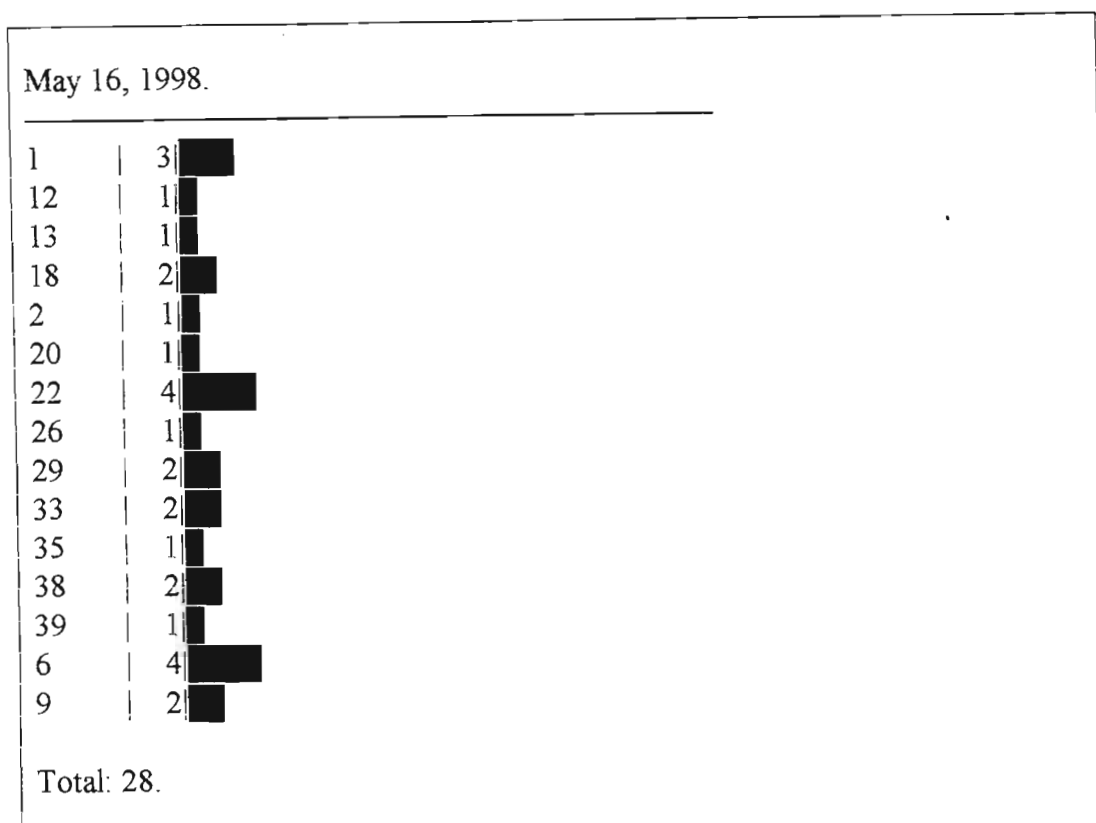
Figure 36: Distribution display of "slavery" query (poem numbers in left column)

Is there evidence here to support Voss' assumptions? A full answer would be possible only through a detailed examination of all the contexts revealed by the analysis, and by further queries suggested by the recursive process of identifying new lines of investigation from the results generated by earlier analyses (an essential part of concordance-based literary analysis - McCarty (1996) stresses this element as the most fundamental in any truly scholarly approach to computer-based analysis). Furthermore, in his article Voss (1990) bases his view on a wider selection of contemporary texts, including court reports of the evidence of slaves, making the inclusion of these additional texts essential to a valid evaluation of his assumptions. It is hoped that this study will be developed in a later paper, but even with the limited contexts provided above, it can be seen that Pringle's use of "slave", "servile" and "servitude" refers not only to "slaves" themselves but also to colonial settlers, revealing that

the texts at the very least transcend simplistic distinctions between European and indigenous inhabitant, or between master and slave.

There is sufficient evidence, even in this superficial display of the "word in context", to explore major differences in the critical interpretation of Pringle's work. For instance, Chapman has characterised Pringle as representing a radical and anti-colonial position whereas Klopper (1990: 57) has argued that his discourse supports fundamental colonialist beliefs.

These contradictory standpoints could be tested by an exhaustive examination of representative words used by Pringle in varying works and contexts. Recursive refining of concordances based on electronic versions of his work could test the extent to which Pringle's discourse does indeed reflect a fundamental adherence to racial superiority, paternalism and a belief in the economic benefits of colonial expansionism on one hand, or the inherent inferiority of indigenous African peoples on the other. One line of departure would be to identify and isolate signifiers in Pringle's discourse (poetic, narrative and in other writings) that might reveal unconscious colonial ideology. Despite ample superficial evidence in his work that he admired indigenous people, actively opposed their exploitation and was sympathetic to interracial marriage, such an analysis could be used to test the conclusions drawn by Voss (1982: 20) and by Klopper (1990: 42) that Pringle's poetry fails to challenge colonial beliefs. For instance, Klopper has already argued convincingly that Pringle's use of the pastoral validated an idealised view of rural harmony, often raised to the level of spiritual sanction, that privileged European values over those of the San or the Xhosa, even when Pringle was attempting to attack Colonial attitudes by adopting the viewpoint of indigenous peoples: e.g. in "Song of the Wild Bushman" (1825) (Klopper 1990: 43). Similarly, Voss has argued that in "Afar in the Desert" Pringle unconsciously links his "firelock" and the presence of God (Voss 1982: 20) revealing a startlingly deep-

seated colonial outlook in the context of a poem in which he elsewhere explicitly rejects colonial society with its "scenes of oppression, corruption, and strife".

Claims made by critics who have used computerised methods for examining semiotic meaning across multiple texts, support the idea that electronic analysis can "shatter the evident surface structures of the text". Similar insights might be gained by subjecting all the available writing by Pringle in electronic form to computerised analysis, and by going on to create a concordance that registered every word in these texts, selecting those which would be most likely to convey ideas of (1) colonial superiority and (2) the barbarism of black people.

By examining the occurrence of such words in the immediate contexts in which they appear, further substantiation for the semiotic interpretations of Klopper and Voss could be found, or indeed, refuted. The nature of this analysis could also be expected to reveal other, unanticipated associations that might suggest further unconscious assumptions present in Pringle's work, relating it to discursive formations of colonial ideology in other contemporary texts. Amongst these could be racial superiority and the idea of the noble savage.

As Olsen (1993) has noted, computer analysis of texts lends itself to the semiotic study of texts. Because a computer program can be instructed to recognise patterns of words or groups of words and the frequency of word occurrence while simultaneously tagging such patterns in the electronic text, the researcher may automate the time-consuming counting and noting of signifiers, and label the semiotic patterns in a specially prepared "marked-up" version of the text, leaving more time for analysing their significance and for interpretation. In the following discussion, TACT and other computer-based programs are applied to Zakes Mda's play *The Hill* (1995) to produce a "dramatological score" based on Keir Elam's methodology for revealing the semiotic structures of drama (Elam 1980).

Elam's eighteen-column scheme produces a diagram that can be used to distinguish patterns that run throughout the dramatic passage and to identify strategies of the speakers as well as the interactions between them. Although Elam admits that such an analysis does not address the levels of *szujet* (the plot) and *fabula* (the basic story line) which would be necessary for interpretation of the play, it does provide a valuable focus on the discourse level, which he feels has been neglected in the more traditional approaches to the study of the language, action, character and interrelationships and the construction of the fictional world of the drama. Fundamental to Elam's approach is deixis - verbal "pointing" indicated by personal pronouns: I, you, we, etc. and by those indicating spatial and temporal dimensions: here, now, there, etc. Dramatic representation, then, is based on the spectator accepting the fictional "here and now" of the characters of the play (through the actors), even though they know this world to be hypothetical. Segmentation of the text proceeds according to a division of the speaker's utterances into "deictic units" shifting the perspective of the discourse amongst these dimensions.

Elam quotes Serpieri on the crucial role in dramatic discourse of deixis - an *I* addresssing a *you, here* and *now* - which allows the dramatic context to be referred to as an "actual" and dynamic world already in progress, :

> In the theatre ... meaning is entrusted in primis to the deixis, which regulates the articulation of the speech acts. Even rhetoric, like syntax, grammar, etc.. are dependent, in the theatre, on the deixis, which subsumes and unites the meaning borne by the images, by the various genres of language (prose, poetry), by the various linguistic modes of the characters, by intonation, by rhythm, by proxemic relations, by the kinesics of the movements, etc. (Serpieri 1978: 20)

In Mda's play, 3000 of the 15000 words used are deictic. In other words, twenty percent of the words spoken by the actors consist of deictic (or pronominal) markers: personal and demonstrative pronouns ("I", "we", "you", "this", "that" etc.) and deictic adverbials ("here", "now", "then" and "there"). Serpieri has cited the high frequency of indexical words in *Hamlet* (a particularly "conceptual" play) to illustrate that drama is considerably more dense in "shifters" (Jakobsen's term for "empty verbal index" - Elam 1980: 139) than poetic or narrative texts. An examination of Pringle's *African Sketches* (1834) confirms Serpieri's observation: in *Poems Illustrative of South Africa*, only ten percent of the words are deictic (1448 of a total of 13836) while in *Narrative of a Residence in South Africa*, the figure is nine percent (12736 of a total of 141626).

In the following exchange from *The Hill*, the degree of deictic density can be clearly seen (my italics):

> VETERAN (incredulously): *You*? They love *you*?
>
> 1st WOMAN: Yes. *We* love *him*.
>
> 3rd WOMAN: *He* is sleeping *here* with *us* tonight.
>
> MAN: *You* are going to be robbed, child of *my* mother. *They* will leave *you* naked like this man.
>
> VETERAN: *They* love *his* money, not *him*.
>
> 2nd WOMAN: *I* didn't know *he* had any money.
>
> MAN: It's really unbelievable *they* should like *him*.
>
> YOUNG MAN (proudly): *I* am a man, child of *my* mother.
>
> VETERAN (incredulously): A man? *He* is a man?
>
> MAN: *You* will *never* understand women. *They* work in mysterious ways.
>
> VETERAN (laughing his booming laughter): *He* is a man?
>
> YOUNG MAN: *You* can laugh as much as *you* like. *That* won't change the fact that *you* are naked. *They* haven't taken *my* pants. *I* don't pay *them*. *They* pay *me*.
>
> MAN: *I* don't understand *this*, child of *my* mother. *They* pay *you*?
>
> 1st WOMAN: Do *you* have to interrogate *him*? (To YOUNG MAN) Remember what *I* told *you*. Shut *your* mouth.
>
> VETERAN: Okay. *I* accept that *they* like *your* manliness. Command *them* to give *me* *my* belongings.

2nd WOMAN: Are *you still* raving about *that*?

3rd WOMAN: Remember what *we* promised to do to *you*. Cut *your* balls.

VETERAN: All right. All right. *You* don't have to use violence.

(Mda 1995: Scene 3)

The value of using the deictic unit as the primary element or category in the dramatological score is the opportunity it provides for the deconstruction and detailed examination of the patterns of interaction amongst characters combining the taxonomies of several different semioticians (Searle, Serpieri, Russell- Elam 1980: 184-191), such as the deictic orientation, the channel of communication, the topic and object of discourse, the illocutionary force and perlocutory effect.

To produce a schematic version of a drama text in which each consecutive segment reflects a new deictic-illocutionary unit (not neccessarily coinciding with specific lines or speeches) the TACT program was used to tag deixis in the text of *The Hill*. The advantages of speeding up the process of marking these segments that form the first category in the dramatological score are considerable. Accuracy of identification is enhanced because TACT tags every one of the deictic units specified by the user. The search list was prepared on the basis of prominal markers identified by Elam and others. The following list, used to tag *The Hill* text, may be used on other texts (the calculation of the "deictic density" of Pringle's texts on page 142 above was performed with the help of this list).

you	585	here	51	that's	10	later	2
I	398	now	48	yourself	10	nowadays	2
my	244	them	48	mea	9	one's	2
we	173	this	48	himself	7	ours	2
it	169	who	46	next	7	ourselves	2
he	162	her	34	whilst	6	themselves	2
that	143	never	27	each	5	whoever	2

Figure 37: Deictic Tags used on *The Hill* text (word frequency in the second column)

Running the TACT tagging program returned the following results (extract from Scene Three shown here). The tagged deictic-illocutionary markers are italicised.

YOUNG MAN: *Someone* is in *our* house.

MAN: What do *you* want in *our* house?

VETERAN: *I* am a veteran miner *just* come from the land of gold.

MAN: That is very good of *you*. But what do *you* want in *our* house?

VETERAN: *This* is God's hill. It is not *your* house.

MAN: *We* live *here* on God's hill. It *is* therefore *our* house.

YOUNG MAN: (inspecting VETERAN): *He* is naked except for a shirt.

MAN: (pointing at the suitcase): *He* is saving the creases of his suits so that when *he* gets home to *his* wife and children *he* should look beautiful.

YOUNG MAN: Look how big *his* suitcase is. *It* must be full of beautiful things for *his* family.

MAN: *I* can imagine the joy of *his* wife when *he* gets home.

YOUNG MAN: Tears of happiness will run in streams.

MAN: And the children jumping about with excitement. '*Ntate o fihlile, ntate o fihlile* - father has come, father has come.'

VETERAN opens his suitcase and displays its emptiness to one and all, as if he has performed a magic trick, laughing in his booming voice.

YOUNG MAN: *It* is empty! Is *he* not from the land of gold?

VETERAN: What does this tell *you*? (Showing them a bracelet around his wrist)

MAN: *It* is the plastic bracelet on which *your* number is engraved.

VETERAN: (enjoying the suspense which he believes he has created): So?

MAN: So *you* are from the land of gold.

VETERAN: That's right. *I* am from the place where men are but numbers engraved on plastic bracelets.

MAN: What do *you* have to show for it?

VETERAN: My plastic bracelet.

MAN: And an empty suitcase. What happened to all the things *you* worked for?

VETERAN: Gone. (He laughs again in his booming voice) Gone the way of all flesh. (Mda 1995: Scene 3)

The tagged extract shown above prepares the text for segmentation as the first category in the dramatological score. A treatment of the first seven segments can be seen in Figure 41 at the end of this chapter.

It can be seen that using a program like TACT for the purpose of semiotic analysis has distinct advantages but also some clear limitations. The advantages include the ease and speed with which the researcher is able to tag the text according to lists of words and patterns determined within the theoretical framework of the study. Once search criteria (like the deictic-illocutionary markers, above) have been saved within the program they may be used again and again on different texts, and as such constitute a handy "tools" for the researcher, turning what would be major time-consuming activities into relatively routine tasks. Yet the program is incapable of identifying *all* the features that the researcher may require. Even deictic-illocutionary markers, relatively discrete and unambiguous elements of the text, pose some problems in this regard. Take the last two lines of the above extract from *The Hill*, for example. There are at least two *implicit* markers:

"And an empty suitcase." - implies (and the audience understands) "*You* have an empty suitcase."

"Gone" - implies "*They* have gone" or even "*I* have lost *them*"

Given the importance that the suitcase and its missing contents have as "objects of discourse" and their significance to the "topic of discourse" (here, the ironic "emptiness" of the material goods acquired through the Veteran's participation in the greater world of the South African gold mining system so central to the other characters' motivations and actions in the play) the "implicit" deictic markers are essential for a valid segmentation of the passage. TACT, in short, cannot be more than a single (though powerful) accessory in the researcher's total analytic project. The relegation of the computer to the position of a

useful adjunct to literary research, rather than its elevation to a self-sufficient methodology is a recurrent theme in studies undertaken with the aid of computer programs:

> What we have to work with here are simple tools - but applied in sophisticated ways. ... [We should use computer programs] with a certain amount of manual intervention, as components in a larger process. Many programs become useful once we realize that *the norm is not one program that will do everything but a number of programs each of which can do something for us* (McCarty 1996:1).

For the analysis presented as Figure 41 (the dramatological score - see the end of the chapter), TACT could be used only as far as the tagging and display of deixis, and the segments then cut and pasted into a spreadsheet program where the rest of the analysis could more easily be performed within a column and row environment. From this point, apart from the fact that a computer is used for layout purposes (rather than using a pen and ink), the identification of each of the dimensions of the dramatological score (diectic orientation, tense, etc.) could be done manually. However, TACT has the potential to semi-automate the process through a number of subsequent steps, including the tagging of additional dimensions. The dimensions of Tense, Channel, Illocutionary force, Explicit performative, and Modality/propositional attitudes lend themselves fairly readily to explicit tagging (although with the same limitations identified in the marking of deixis as described above). At the end of the chapter, the first seven segments are presented in dramatological score format in Figure 41, followed by a table, adapted from Elam (1980) explaining how the columns should be read.

However, before the dramatological score can be drawn up, the text must be prepared further. A two-stage procedure is essential here as the tagging of the above dimensions can be undertaken only after the deictic segments have themselves been tagged, and a "new" version of the text prepared, arranged in deictic segments corresponding to "Column 2" of the dramatological score, each segment associated with the corresponding speaker (Column

3). To reach this point (Stage 1: Preparation of the deictically-ordered text), the following steps are necessary:

Step 1: Using the Group of deictic/illocutionary markers (Figure 37, above) create a Query file within the textbase, and invoke a KWIC display. After modifying the display to convert the "headword" to a ">" and to order the text sequentially, export the display to a text file.

Step 2: Open the text file in a word processor (See Figure 38 below).

```
478   quickly shuts his suitcase.} | «YMAN» >Someone is in our house. |
 478   his suitcase.} | «YMAN» Someone is in >our house. | «MAN» What do you
 479        is in our house. | «MAN» What do >you want in our house? | «VET»
 479      house. | «MAN» What do you want in >our house? | «VET» I am a
 480      do you want in our house? | «VET» >I am a veteran miner just come
 481    come from the land of gold. | «MAN» >That is very good of you. But
 481    of gold. | «MAN» That is very good of >you. But what do you want in
 481   That is very good of you. But what do >you want in our house? | «VET»
 481    good of you. But what do you want in >our house? | «VET» This is
 482      do you want in our house? | «VET» >This is God's hill. It is not
 482     house? | «VET» This is God's hill. >It is not your house. | «MAN»
 482   | «VET» This is God's hill. It is not >your house. | «MAN» We live
 483   hill. It is not your house. | «MAN» >We live here on God's hill. It
 483   It is not your house. | «MAN» We live >here on God's hill. It is
 483   | «MAN» We live here on God's hill. >It is therefore our house. |
 483   here on God's hill. It is therefore >our house. | «YMAN»
 484       | «YMAN»  {inspecting VETERAN} >He is naked except for █ | a
 485   | «MAN» {pointing at the suitcase} >He is saving the █ | creases
 485      He is saving the █ | creases of >his suits | so that when he
 486      the █ | creases of his suits | so >that when he gets home to his
 486   | creases of his suits | so that when >he gets home to his wife and
 486   suits | so that when he gets home to >his wife and children he
 486   he gets home to his wife and children >he should look beautiful. |
 487   look beautiful. | «YMAN» Look how big >his suitcase is. It must be
 487   «YMAN» Look how big his suitcase is. >It must be full of beautiful
 487    be full of beautiful things █ | for >his family. | «MAN» I can
 488      things █ | for his family. | «MAN» >I can imagine the joy of his
 488      | «MAN» I can imagine the joy of >his wife when he gets home. |
 488   can imagine the joy of his wife when >he gets home. | «YMAN» Tears
 492    father has come.' | {VETERAN opens >his suitcase and displays its
 492      opens his suitcase and displays >its emptiness to one █ | and
 492    emptiness to one █ | and all, as if >he | has performed a magic
 493   performed a magic trick, laughing in >his booming voice.} | «YMAN»
 494       in his booming voice.} | «YMAN» >It is empty! Is he not from
 494      voice.} | «YMAN» It is empty! Is >he not from the land of gold?
 495   the land of gold? | «VET» What does >this tell you? {Showing them
 495   of gold? | «VET» What does this tell >you? {Showing them a bracelet
 495   What does this tell you? {Showing >them a bracelet █ | around his
 495   {Showing them a bracelet █ | around >his wrist} {enjoying the
 497   them a bracelet █ | around his wrist} >{enjoying the suspense which
 497   them a bracelet █ | around his wrist} >{enjoying the suspense which
 498   wrist} {enjoying the suspense which >he believes he has created}█ |
 498       the suspense which he believes >he has created}█ | So? |
 499      he has created}█ | So? | «MAN» So >you are from the land of
```

Figure 38: Modified KWIC display

Step 3: After adding a tab separator before each ">", convert the text to a table.

Step 4: Import the table into a spreadsheet (which has more versatility in managing the

478	quickly shuts his suitcase.} | YMAN >Someone is in our house. |	>Someone is in our house.	YMAN
478	his suitcase.} | YMAN Someone is in >our house. | MAN What do you		
479	is in our house. | MAN What do >you want in our house? | VET	What do >you want in our house?	MAN
479	house. | MAN What do you want in >our house? | VET I am a		
480	do you want in our house? | VET >I am a veteran miner just come	am a veteran miner just come the land of gold.	VET
481	come from the land of gold. | MAN >That is very good of you. But	>That is very good of you	MAN
481	of gold. | MAN That is very good of >you. But what do you want in		
481	That is very good of you. But what do >you want in our house? | VET	But what do >you want in our house?	MAN
481	good of you. But what do you want in >our house? | VET This is		
482	do you want in our house? | VET >This is God's hill. It is not	>This is God's hill.	VET
482	house? | VET This is God's hill. >It is not your house. | MAN	>It is not your house	VET
482	| VET This is God's hill. It is not >your house. | MAN We live		
483	hill. It is not your house. | MAN >We live here on God's hill. It	>We live here on God's hill.	MAN
483	It is not your house. | MAN We live >here on God's hill. It is		
483	| MAN We live here on God's hill. >It is therefore our house. |	>It is therefore our house	MAN
483	here on God's hill. It is therefore >our house. | YMAN		
484	| YMAN {inspecting VETERAN} >He is naked except for | a	>He is naked except for	YMAN
485	| MAN {pointing at the suitcase} >He is saving the | creases	>He is saving the | creases of his suits | so >that when he gets home to his wife and children >he should look beautiful.	MAN
485	He is saving the | creases of >his suits | so that when he		
486	the | creases of his suits | so >that when he gets home to his		
486	| creases of his suits | so that when >he gets home to his wife and		
486	suits | so that when he gets home to >his wife and children he		
486	he gets home to his wife and children >he should look beautiful. |		
487	look beautiful. | YMAN Look how big >his suitcase is. It must be	Look how big >his suitcase is.	YMAN
487	YMAN Look how big his suitcase is. >It must be full of beautiful	>It must be full of beautiful things for >his family.	YMAN
487	be full of beautiful things | for >his family. | MAN I can		
488	things | for his family. | MAN >I can imagine the joy of his	>I can imagine the joy of his wife when >he gets home	MAN
488	| MAN I can imagine the joy of >his wife when he gets home. |		
488	can imagine the joy of his wife when >he gets home. | YMAN Tears	Tears of happiness will run in streams	YMAN
492	father has come.' | {VETERAN opens >his suitcase and displays its	Father has come..	MAN
492	opens his suitcase and displays >its emptiness to one | and		
492	emptiness to one | and all, as if >he | has performed a magic		
493	performed a magic trick, laughing in >his booming voice.} | YMAN		
494	in his booming voice.} | YMAN >It is empty! Is he not from	>It is empty!	YMAN
494	voice.} | YMAN It is empty! Is >he not from the land of gold?	s >he not from the land of gold?	YMAN
495	the land of gold? | VET What does >this tell you? {Showing them	What does >this tell you?	VET
495	of gold? | VET What does this tell >you? {Showing them a bracelet		
495	What does this tell you? {Showing >them a bracelet | around his		
495	{Showing them a bracelet | around >his wrist} {enjoying the		
497	them a bracelet | around his wrist} >{enjoying the suspense which		
497	them a bracelet | around his wrist} >{enjoying the suspense which		
498	wrist} {enjoying the suspense which >he believes he has created} |	>It is the plastic bracelet on which your number is engraved	MAN
498	the suspense which he believes >he has created} | So? |	So?	VET
499	he has created} | So? | MAN So >you are from the land of	So >you are from the land of gold	MAN

Figure 39: Deictic/illocutionary segments:Initial spreadsheet layout

Step 4: Import the table into a spreadsheet (which has more versatility in managing the column and row framework required for the dramatological score).

1	2	3
	Segment	Speaker
1	>Someone is in our house.	YMAN
2	What do >you want in our house?	MAN
3	>I am a veteran miner just come the land of gold.	VET
4	>That is very good of you	MAN
5	But what do >you want in our house?	MAN
6	>This is God's hill.	VET
7	>It is not your house	VET
8	>We live here on God's hill.	MAN
9	>It is therefore our house	MAN
10	>He is naked except for a shirt	YMAN
11	>He is saving the | creases of his suits | so >that when he gets home to his wife and children >he should look beautiful.	MAN
12	Look how big >his suitcase is.	YMAN
13	>It must be full of beautiful things for >his family.	YMAN
14	>I can imagine the joy of his wife when >he gets home	MAN
15	Tears of happiness will run in streams	YMAN
16	And the children jumping about in excitement.	MAN
17	father has come, father has come.	MAN
18	>It is empty!	YMAN
19	Is >he not from the land of gold?	YMAN
20	What does >this tell you?	VET
21	>It is the plastic bracelet on which your number is engraved	MAN
22	So >you are from the land of gold	MAN
23	>That's right.	VET
24	>I am from the place where men are but numbers engraved on plastic bracelets	VET
25	What do >you have to show for it?	MAN
26	>My plastic bracelet.	VET
27	What happened to all the things >you worked for?	MAN
28	Gone. Gone the way | of all flesh.	VET

Figure 40: Deictic/illocutionary segments: Edited spreadsheet layout

Step 5: Using the tagged text in the first column, cut and paste each deictic/illocutionary segment into the second column and enter the Speaker's name into the third. Delete the first column and any empty rows. Add a new Column 1 to number the segments and add headers for each column.

Step 6: Further edit the segments, adding words and possibly implicit deixis that may have been omitted from the original KWIC display.

The procedure described here, although not fully automated, considerably speeds up the process of preparing the basis of the dramatological score with the deictic segment as the basic organisational unit.

Having completed Stage A in this manner, the researcher may put aside the original tagged text, and use TACT again to mark up the newly-segmented text. In the preparation of the new textbase, the tagging of the explicit dimensions (Tense, Channel, etc.) may commence on the orderly basis provided by the segmented version of the play.

Any attempt to go on to a third stage - the tagging of the more semantic dimensions such as Topic/object of discourse or Lexemes/isotopies/semantic paradigms - would possibly be self-defeating as the researcher would be more likely to identify the significant elements within these dimensions retrospectively. It would therefore be more efficient to add these columns to the dramatological score manually: it would be well to heed McCarty's caveat here - the computer is a useful tool only to the extent that its capabilities contribute to the researcher's purpose, not limit or obstruct it (McCarty 1996: 1).

The following figure and table, adapted from Elam (1980), illustrate what can be achieved by this process.

Dramatological Score (Elam) Mda's The Hill

1	2 Seg.	3 Spe.	4 List.	5 D.O.	6 T.	7 Ch.	8 Topic/object	9 I.F.	10 E.P.	11 P.E.	12 Implic./rhet	13 Mod.	14 An.	15 Met.	16 Other functs.	17 Lexs./isotops	18 Cultural codes
23	Someone is in our house	Yman	Man (Vet)		Pres.		T: Homeless O: Vet	D		+	Impl: disapproval	p	'someone'			ownership	ownership
24	What do you want in our house?	Man	Vet		Pres.		T: Homeless O: Vet	I?	+	−	Impl: "get out"	Op	'Vet'			ownership	
25	I am a veteran miner just come from the land of gold.	Vet	Man (Yman)		Pres.		T: Migr. labour O: Vet	D			Impl: superiority	p	'Vet'			migratory labour	
26	That is very good of you.	Man	Vet		Pres.		T: Relations O: Vet	E			Impl: sarcasm	p	'Vet'				
27	But what do you want in our house?	Man	Vet		Pres.		T: Homeless O: Vet	I?	+		Impl: "get out"	?p	'Vet'			ownership	
28	This is God's hill.	Vet	Man (Yman)		Pres.		T: Aspir/Hope O: Hill				Impl: "you do not own this space"	Kp	'God'			God/religion	spirituality
29	It is not your house.	Vet	Man (Yman)		Pres.		T: Homeless O: Vet	−			Impl: "get out"	~p	'Vet'			ownership	

"The score can be read horizontally, so as to determine the different levels at which a given segment works simultaneously, or vertically, in order to establish the patterns running throughout the passage, the strategies of the given speaker, the development of the interaction and so on." (Elam 1980: 185)

Figure 41: Dramatological Score. See the table (below) for an explanation of the columns.

An explanation of the columns in the Dramatological Score (Figure 41).	
1. Line	These line references were generated in TACT and based on the scanned copy of the play (1995).
2. Segment	The segments, as can be seen from the above discussion, do not necessarily coincide with a new line or speech; they register new deictic or illocutionary units.
3. Listener (s)	The first name is that of the actual addressee; names in parentheses indicate other characters present, presumably listening but not directly addressed.
4. Deictic orientation	Explicit deictic references are represented by the following symbols in **Figure 41** (for the full set, see Elam 1980: 186).

		Proximal deictic orientation towards context (*here*) (Category: Proximal deictic orientation)
		Proximal deictic orientation towards listener (*you*) (Category: Person-deixis)
		Proximal deictic orientation towards speaker (*I*) (Category: Person-deixis)

6. Tense	Orientation towards present, past, future. Symbols are as follows: Pres.: present; P.P.: present perfect; Pas.: past; Fut.: future; Subj.: subjunctive; Imp.: imperative.
7. Channel	Elam points out that this is a rather heterogeneous column including the dominant physical channels along which the characters operate (acoustic or visual), together with the physical, psychological, emotional and ideational states expressed by the speaker, references entailing movements, etc. Only those symbols used in the extract from *The Hill* are given here.

An explanation of the columns in the Dramatological Score (Figure 41).		
	(body figure symbol)	The body (references to the speaker's or interlocutor's physical involvement in the scene).
	(heart symbol)	Emotional state (reference to attitude or reaction).
	(thought cloud symbol)	Cerebration (ideational bias, expression or concepts).
8. Topic/ object of discourse	By "topic of discourse" is meant the global theme or central concern of the exchange. The passage analysed exhibits four main topics: homelessness, the migratory labour system, the hopes and aspirations of the MAN and YOUNG MAN and the relationship between all three characters. See my discussion above for further discussion on this issue. The "object of discourse" is the individual person, object, event or notion referred to within each segment.	
9. Illocutionary force	Elam adapts Searle's taxonomy for this column. Only the symbols used in **Figure 41** are given here (for the full set, see Elam 1980: 188).	
	D	Declarations
	!?	Question
	E ⊃	Greeting/ salute
	⊢ $\underline{\Omega}$	Hypothesis
	⊢ -	Negation
10. Explicit performative	A speech act whose illocutionary force is specified through a performative verb. Indicated by a + sign where such a verb is present.	

An explanation of the columns in the Dramatological Score (Figure 41).	
11. Perlocutionary effect	If a command is obeyed it achieves a perlocutionary effect and is indicated by a + sign. Failure to achieve the effect is indicated by a - sign, and where it rendered null and void, an **X** is used.
12. Implicatures/ rhetorical figures	Elam identifies these as unspoken meanings founded on the co-operative principle, indicating context-bound figures of speech. Implicatures such as sarcasm and superiority are identified in the exchange from *The Hill*.
13. Modality/ propositional attitudes	This is a "logical" column concerned with the attitude expressed by the speaker towards the propositional content of the utterance (belief, possibility, etc.) And thus the logical modality governing the segment. Once again, only the symbols used in the Mda passage are given here.

p	Alethic necessity (it is the case that p)
Op	Deontic modality of obligation (You must p)
?p	Alethic possibility (is p?)
Kp	Epistemic certainty (I know that p)
~p	Alethic necessity (negative) (it is not the case that p)

14. Anaphora	There are two kinds of anaphoric reference:
⇒	Internal reference to an antecedent in the linguistic co-text (given in quotation marks).
⇐	External reference to an extra-linguistic "it" or "he" not present in the current situation and not specified in the co-text.

An explanation of the columns in the Dramatological Score (Figure 41).	
15. Metalanguage	This column includes references to messages, the code, idiolect, the act of speaking or listening, language at large, etc. Language itself does not serve as an object of discourse in this particular extract.
16. Other functions	Any semiotic function not indicated elsewhere.
17. Lexemes/ isotopies/ semantic paradigms	This is a broadly "semantic" column designed to indicate the chief levels of semantic and lexical coherence running throughout the passage. The retrospective nature of the selections for this column are mentioned in my discussion, above. Private ownership, the migratory labour system and the ironies of God/hope and organised religion are identified here.
18. Cultural codes	This column is closely related to the previous one, and indicates the main social, ideological, religious, moral, epistemological and intellectual norms invoked in the dialogue. Ownership and conceptions of spirituality are indicated here.

(Adapted from Elam 1980: 185-191)

In the next chapter, we return firstly to the hypertext environment of the SALIT Web as a whole as embodied in its publication as the Beta 01 version on CD-ROM, and then to a consideration of its place as a subset of a local - and a global - networked set of electronic resources in South African literature.

Chapter 8 The SALIT Web CD-ROM in context

The investigation of research projects, electronic text resources and software applications reflected so far in this study have been applied to the production of the SALIT Web on CD-ROM. The CD-ROM that accompanies this thesis represents the practical output of the study. As an application and adaptation of internal architecture, organisational structures and search features of the many different sources described and evaluated in the course of this study, the SALIT Web is an integral part of the project as a whole. Thus it stands on its own as an artifact embodying the substance of this discussion. For the purposes of documentation and discussion, the SALIT Web has proved to be something of a moving target. Until the production of the "Beta 01" version of the CD-ROM in September 1998[10], the structure and contents changed frequently and some of the experimental features were omitted from Beta 01 but will be incorporated into future versions of the SALIT Web. References here to "the present version" of the SALIT Web are to the September 1998 Beta 01 release that accompanies the thesis.
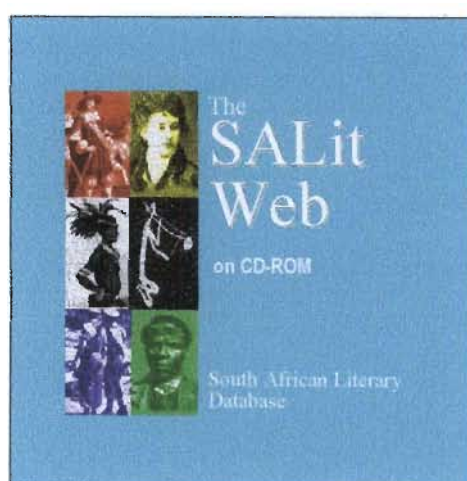

Figure 48 CD-ROM cover

---

[10]   See Appendix G for notes on installing this experimental version.

While The SALIT Web is a stand-alone resource for various learning, teaching and research activities in South African literature, it is also intended as a prototype for the further development of a virtual library of rare and marginalised South African writing. The CD-ROM is any and all of the following: a bibliographic database, a virtual library, a multimedia encyclopaedia, an archive for the preservation of marginalised texts, a textual analysis tool, and a learning and teaching programme. The sheer volume of information now included in the Web - due in no small part to the considerable efforts of both Prof. van Wyk, Head of the CSSALL and Rashmi Jadhunundhan, research assistant at the Centre - has made it impossible in this discussion to trace anything but a few pathways through the available resources. Only by using the CD-ROM itself can the user begin to appreciate the value it may have as a useful research and learning tool.

The purpose of this chapter is to locate the SALIT Web in the larger networked environment that lies beyond the fairly narrow perimeter of the CD-ROM. To begin with, let us examine the SALIT Web and its internally-linked contents within the context of a global networked cultural heritage.

Planning an electronic literary resource had to take into account present developments in electronic publishing: the increasingly common practice of publishing journals, books and encyclopaedias on CD-ROM and the Internet. While allowing the reader to view a publication much as it would appear on a printed page - but now transferred to a computer display - the electronic text may be accompanied by a powerful computerised search apparatus that delivers easy access to cross-references either within the article or to other articles on the disk or network. Quick, comprehensive cross-referencing clearly has advantages for a reader who might require a list of related entries (biographical information, reviews, etc.), or may wish to read and keep a printed record of the selected text. The

SALIT Web exploits the medium by accelerating the type of searching and cross-referencing already familiar to researchers and can even preserve a particular search "route" so that it might be used again later.

Despite the benefits offered by an individual electronic text title, literary scholars demand a whole range of related texts at their fingertips, for quick referencing, cross-referencing and annotation. But at the moment most publishers do not offer the full scholarly potential of the medium because they tend to be unwilling to produce a text that implies sharing of information outside the proprietary "package" of the work itself (see the references to copyright in Chapter 4). Any electronic publication may be a hypertext - with embedded links that enable the user to navigate from one discrete chunk of information to another by activating highlighted text in the original document with a pointing device. The reader may just as readily use this feature to reach out via embedded links in the text to related print, archival, electronic and critical material located elsewhere. All such texts share a common digital medium, that may be searched for references, common words or word patterns, especially if they use a dedicated textual analysis application like TACT (see Chapter 7). Within the bounds of copyright, the SALIT Web does make the texts in its Library section available for searching and analysis, and each text is hyperlinked to bibliographic entries (author, title, date) for easy identification. But embedded links could just as easily permit the reader to move beyond the perimeter of the SALIT Web to remote text databases, archives, libraries and collections anywhere in the world. Hypertext coding (embedded links) within the full texts and in the bibliographic entries could give the user access to a world-wide library of associated electronic texts. Although this potential versatility depends on textually-based hyperlinks and descriptive mark-up (document structure and word

tagging, as in the TEI scheme), the tagging does not obstruct readability because electronic texts can be made to emulate the attractive format and page layout of their print-based counterparts: the Oxford English Dictionary Second Edition is one notable example of the happy co-existence of print and electronic versions. The user of the CD-ROM dictionary may wish to search for the earliest quoted word beginning with "technolog": in response to enquiry the dictionary opens a "window" displaying the appropriate page in the distinctive typeface of the Oxford University Press documenting the 1615 use of the word "technologie" by Buck in *Third University English*, p. xlviii (Seaman 1994: 16-17). Using the SALIT Web, users would be able to access information like this from their desktops. As an electronic artefact, the SALIT Web is a type of meta-text incorporating a hierarchy of
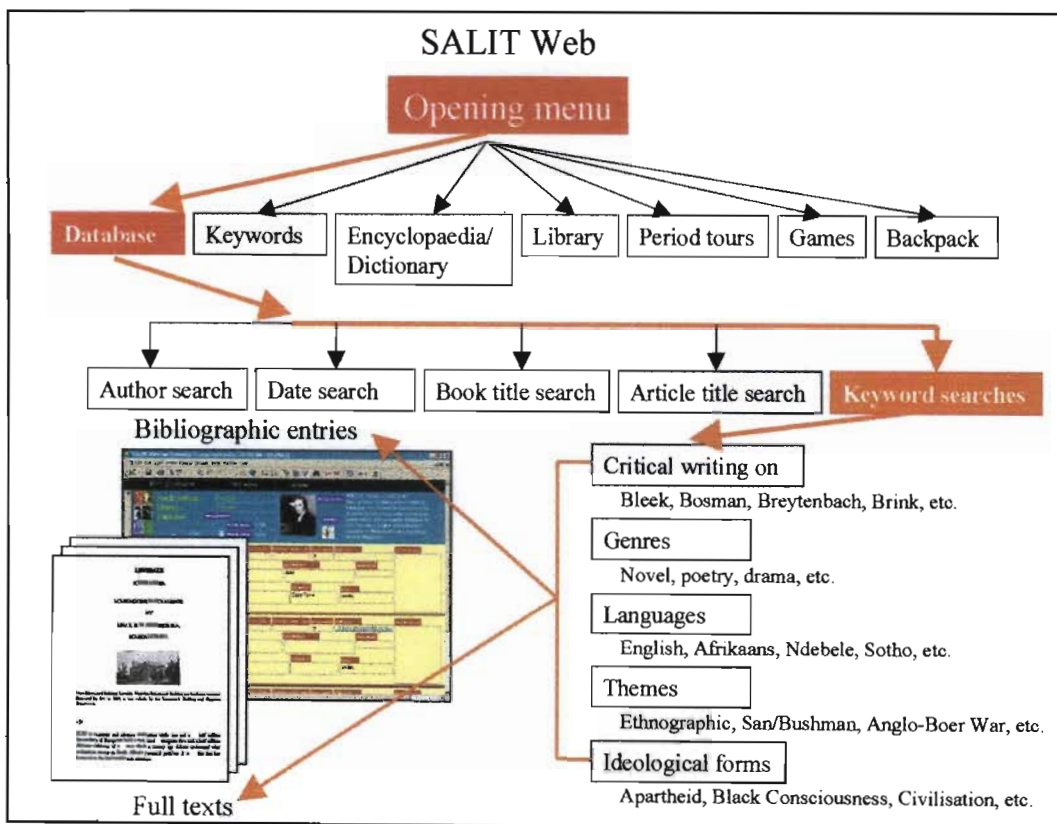

Figure 49: Overall structure of the SALIT Web

linked material: primary entries including literary biographies, a chronological display of titles, historical periods, discursive and thematic formations, etc. that in turn contain cross-

references, bibliographic references and full texts. A hypertext structure is non-sequential, and although the SALIT Web material is hierarchically organised, progress through the material is user-driven.

Many scholarly texts, including encyclopaedias, already have a hypertextual structure (see Chapter 3). As has been discussed, any table of contents, index, glossary, footnote or bibliographic reference implies reader movement (or "navigation") amongst several texts both within the printed document and outside it. The reader of a scholarly document may be prompted by a reference within the text to leave the main text to read a footnote providing additional commentary or information about sources, biographical or historical background. This, in turn, may lead the reader to another text (such as a critical article or a contemporary history) "outside" the original document itself. Conventionally, such a step would mean accessing the referenced document through a library. The reader might then return to the main text again. This type of reading may be characterised as "constitut[ing] a mental model of hypertext" (Landow & Delaney 1991: 5).

However, in a print-based environment (i.e. reading a book) even though navigation through associated texts is central to the reader's activity, the process can be protracted "because the references (or linked) materials lie spatially distant from the reference mark" (Landow & Delaney 1991: 5). If one were looking for the rare or marginalised South African texts included in the SALIT Web Library, the difficulties of access to a printed copy could extinguish the process completely. The Oxford English Dictionary search illustrates the ease with which documents may be accessed in a hypertext network. Were the search to have been performed in a full-text database the scholar could have pinpointed and retrieved the full text of the novel, play, poem or article to study the context in which the word was used by all authors whose work is stored in the literary corpus. In Chapter 1, I described how Wolff used the ARTFL (Project for American and French Research on the Treasury of

the French Language) database, to test his assumptions about the use of the words "real"
and "ideal" in nineteenth century French prose:

> In order to pursue the question of realism and idealism in the nineteenth century, one
> could begin to look for presuppositions about the about the real and ideal in texts from this
> period. [The database lists] words that often function to introduce presuppositions: *quand,
> moment, quant, lorsque*, etc. The keywords in this list occur in sentences that contain
> words matching the patterns. (Wolff 1994: 41).

Further examples of textual analysis can be found in Chapter 7 using the TACT textual
analysis program. If we now re-focus our view of the SALIT Web, the overview diagram in
Figure 50 could highlight the potential offered by the Library texts for textual analysis:
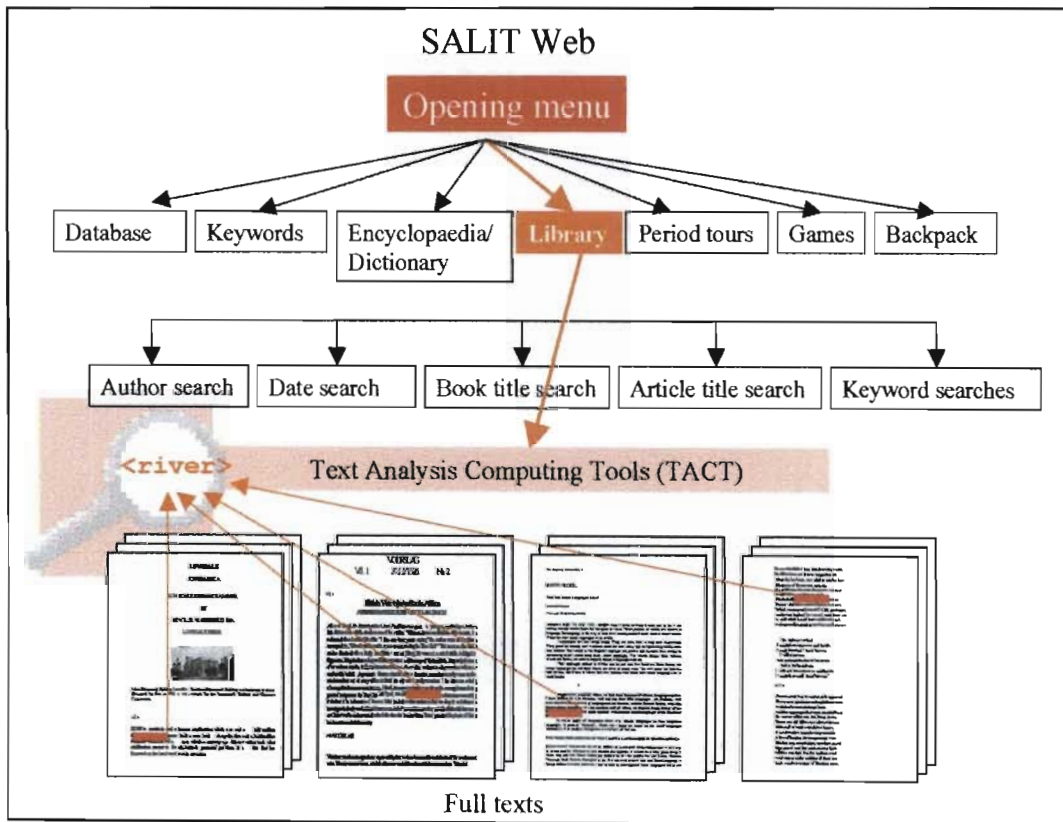
Figure 50: SALIT Web: textual analysis focus

In Figure 50, a search for the word "river" (incidentally the highest frequency "place" designation word in Pringle's *Narrative of a Residence in South Africa*") immediately displays the use of the word in several South African texts. In his account of using the ARTFL database, Wolff goes on to warn that it would be impossible to map any discursive space definitively, but electronic text resources allow the scholar to explore intertextuality in a manner impossible before the advent of computers. The illustrations of textual analysis performed on the Pringle and Mda texts in Chapter 7 bear out Wolff's comments - the tools available for the analysis of digital text can be very helpful to the researcher, but they can play only a mechanical part in the analytic process. By generating re-orderings of the original text according to semiotic categories determined by the researcher, linguistic and literary structures can be studied more quickly and more comprehensively, but the tasks of synthesis and interpretation cannot be delegated to the computer.

If the hierarchy depicted in Figure 49 is extended to trace the associative direction

Figure 51: From SALIT Web to SABINET to full text

suggested by, for example, "Bibliographic Entries", a reader either directly or indirectly linked via the computer to the South African Bibliographic Network (SABINET) the diagram could be extended.
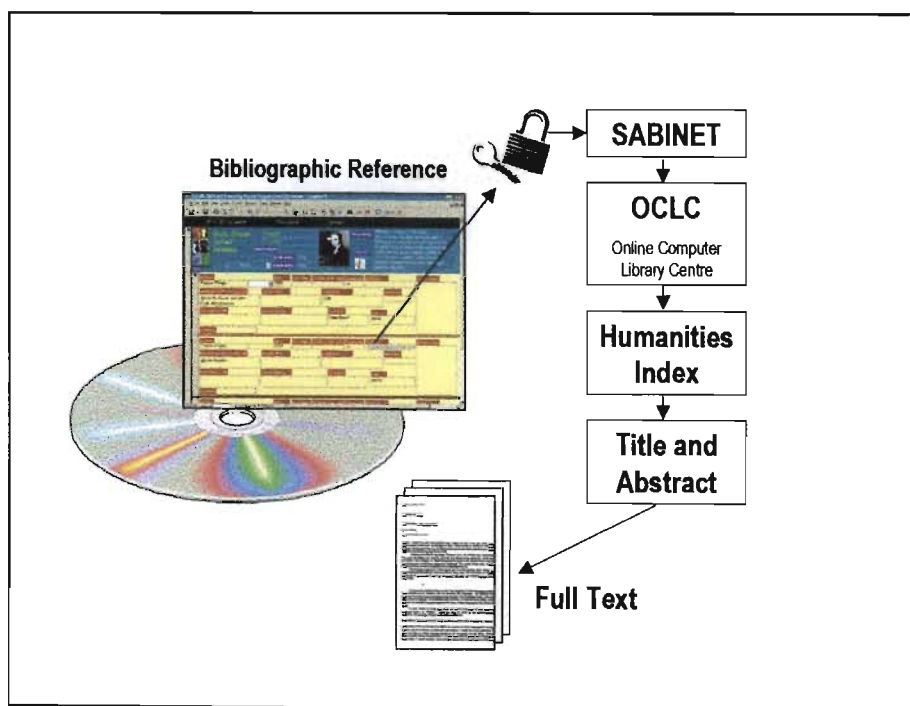
In the example shown in Figure 51, the reader is prompted by a pointer in CD-ROM to an electronic access route - SABINET and then OCLC - through to the designated article. But if the electronic (CD-ROM) version of the encyclopaedia were encoded for the purpose, it is entirely feasible that an appropriate hypertext link embedded in the bibliographic entry would perform the same activity automatically, following each of the steps in Figure 51 and culminating in the appearance of the full text of the referenced article directly onto the reader's computer screen. Internet users, for instance, perform similar searches as a matter of course. In the case of the SALIT Web, the reader could then read the text, and by using the built-in search facilities, either navigate through the retrieved text to the page included in the citation, or failing that perform a word, phrase or a "keyword-in-context" search (see Wolff, above) to target the desired content (a sample page display from a SABINET search can be seen in Chapter 2).

It must be stressed that the automated search activity described here is by no means whimsical or far-fetched. For the last five years undergraduate students at the University of Virginia have been able to query extensive text corpora in just this way from computers located in their residence dormitories (Seaman 1993). The development of electronic text centres in the humanities has made available both the software and the texts to perform automated searches with flexibility and simplicity. The SALIT Web strives to offer the same access to South African texts.

In the light of the research into electronic texts mentioned here, planning a resource such as the SALIT Web without considering its extensibility into a network of associated electronic sources would be as bizarre as proposing that the work be hand-written on vellum. The

examination of the current status of electronic encyclopaedias presented in Chapter 3 suggests that the *South African Literary Encyclopaedia*, out of which the SALIT Web project grew, is likely to become a CD-ROM publication itself. A digital version of the encyclopaedia would ensure that the scholarly activity invested in its production were optimised by utilising advanced computing tools. By taking the electronic route, the *Encyclopaedia* would avoid being doomed to an anachronistic mute existence, cocooned in an envelope of ink-text too costly for most South Africans to afford, or worse, locked into an incompatible publisher-specific encrypted code, capable of effecting only page-layout and formatting (procedural mark-up as opposed to descriptive mark-up -see Chapter 6).

How then would one position the encyclopaedia within the larger realm of electronic publishing - the "docuverse" (Landow and Delaney 1991) or the "Information Omniverse" (Seaman 1994), to use hypermedia jargon? The following diagram is an attempt to provide a perspective that defines the parameters of the specific SALIT Web "package", while locating it within the wider network of existing (and projected) electronic text resources:
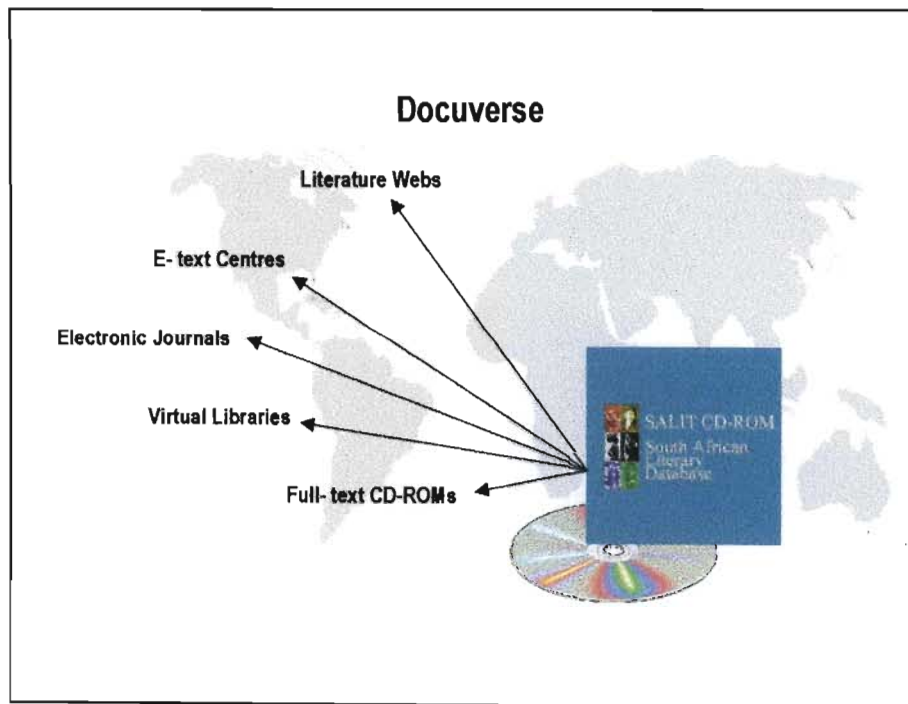

Figure 52: The SALIT Web and the Docuverse

The above model allows the SALIT Web as a distinct identity as a publication within the docuverse, while leaving links open to external associations. The hypertext possibilities differ only in terms of scale and physical constraints: the maximum capacity of one CD-ROM is approximately 600 megabytes, while remote e-text collections offer almost limitless storage capacity. Both intertextual and intratextual analysis of texts, as described earlier, are possible. A user may search for cross-referenced texts and record the results, or search for individual words using software - Folio Views or TACT - equipped with a KWIC "keyword-in-context" feature. Both the SALIT Web package and the wider networks shown in Figure 52 allow for the close analysis of individual or large corpora of texts (e.g. TACT referred to above). One should imagine two distinct (though related) levels of activity: searching for relevant commentaries, critical articles and archive materials across a wide range of texts as part of a literature study, or the analysis of a specific text to test text-critical assumptions. These "levels" apply both within the SALIT Web and the wider network, but the latter implies the existence of a body of full texts in electronic form that far exceeds the present collection in the SALIT Web Library. Ideally, the SALIT Web should be hyperlinked beyond its immediate domain to other collections of electronic text via the Internet. Even though from an electronic point of view, these would all fit seamlessly into the wider docuverse, for practical purposes they currently inhabit separate domains (the "gateway" between the SALIT Web and SABINET, described above illustrates the point). For copyright reasons alone, it would be necessary at first to keep the SALIT Web behind a password-protected gateway, just as the SABINET resources are. Some subscription and registration options are discussed in Chapter 10.

Documentation of the design and construction of the database, the hypertext environment and the multimedia files can be seen in Chapter 10. Earlier in this chapter, we examined the relationship between the SALIT Web and the rest of the "docuverse". As it stands, the CD-ROM version already has the capability of reaching out via hypertext links to the wider world of the Internet. But to be truly part of the "big picture", other users of the Internet

need to be able to access our resources. In the next chapter, the future development of the SALIT Web on the Internet is discussed.

Chapter 9 The way forward: Internet access

Once the SALIT Web is available on the Internet, two related sites could immediately be integrated into its structure: the existing SALIT e-mail discussion group and an on-line version of the CSSALL's journal, *Alternation*. Quite apart from the rich literary resources of the SALIT Web, a combination of discussion group and journal linked on a common Internet site could meet a number of needs of the academic community in South African literary studies. Careful management of such a site could enable a group of academics to return full circle to the original purpose of the academic journal: a lively, ongoing discussion or seminar radiating from the planned, considered critical viewpoints of its participants (the journal articles). The literature reports experiments with this type of on-line structure that have succeeded in generating discussion around articles published in an electronic journal (e-journal) while maintaining the traditional peer-review process, and the academic "integrity" of the published articles. The milieu in which this publication and discussion would take place is not unlike the ALN environment described in Chapter 5, and could indeed be situated within a distributed application like Lotus Notes. However, the necessity of the participants to have the software would create an unnecessary obstacle, and I believe the e-journal/discussion site would be more likely to flourish through direct Internet access.

I will give a brief account of the existing SALIT discussion group and the capabilities of such groups in general, and then document the progress I was able to make with an electronic version of *Alternation*. Thereafter I will discuss how these may be brought together within the SALIT Website with reference to other publishers of academic journals who have experimented with similar models.

Firstly, what is a discussion group? On-line discussion groups are made up of a number of e-mail users who receive copies of all the messages sent out by other members of the group.

There are server-based application programs that manage this process such as Listserv, on which the current SALIT discussion group operates via the university server. Although a discussion groups can operate merely as bulletin boards, containing announcements to all its subscribers, the very existence of the group tends to promote conversation amongst the members, that often become topical discussions. While the SALIT discussion group is an "open" list that permits anyone with an e-mail address to join, academic groups usually operate as "closed" or even "moderated" groups. A closed group requires the prospective member to be approved by the person managing the list, after which all submissions made by the new member will be posted. In a moderated list, all posts go through a moderator for approval and messages may be rejected or edited. The closed, moderated type of group is the ideal for an academic discussion group linked to an e-journal, as the approval of participants loosely resembles the process of peer-review, and may be adapted for actual reviewing of individual articles. An important caveat is the time-consumed by such moderation, and a large flourishing group would require a substantial commitment from the moderator (or moderators). Although the SALIT group was deliberately left "open" at the outset in order to be as accommodating to new members as possible and encourage participation, it has succumbed to one of the pitfalls of Internet communication, the so-called "spam" factor, where either commercial or crank members can begin to flood the in-boxes of subscribers with the electronic equivalent of "junk mail". One way of dealing with this is to apply an automatic filter to the list, effectively screening out any messages originating from the offending e-mail address, but in the longer term the most effective answer is moderating the messages.

An examination of the messages posted on SALIT since its inception in 1996 reveals that most of the messages posted have either been announcements (invitation for conference paper submissions, new publications and a few job advertisements) or information seeking questions ("where will I find ..."). Both categories are useful for the participants and serve a worthwhile purpose, but neither constitute academic debate. To initiate debate, a group

needs to respond to a proposition, and this is what gives the e-journal-discussion group combination such potential. Potter (1996) observes that computer-mediated discussion encourages involvement by participants who might otherwise be intimidated by face-to-face encounters in a seminar room. She explains that participants can spend as much time as they like formulating their responses, and if necessary refining them before actually posting them into the discussion. Other benefits of this type of discussion, for example the archiving of contributions and responses, have been mentioned as part of the features of an ALN in Chapter 5.

Now we will turn to e-journals, the other component of the integrated web that I am proposing for SALIT. The trend towards on-line publication of journals has already been mentioned (see Chapter 3) in my discussion of electronic encyclopaedias. The OCLC alone has the full text of over 50 journal titles available over the Internet and many other individual journals are available for individual subscription. Newspapers have also developed rapidly, setting up electronic versions on-line, often providing a searchable archive to subscribers. In South Africa several print-based newspapers have Internet editions, for instance the *Mail & Guardian* which also offers free access to their archive, a useful repository of reviews of South African literature and film. A more far-reaching search of free South African newspaper archives can be performed through the Rhodes University "Gogga" search engine that is linked to several individual newspapers including the *Mail & Guardian*. While this type of on-line journal or newspaper is indicative of a tendency to publish on the Internet, the style of publication is essentially of the linear print-based kind - electronic? yes; but interactive? no. Even the way in which some of the e-journals are made available - in Adobe Acrobat format, for example - reveals a clinging to the printed page format rather than an acceptance of hypertext as a medium of transmission. Acrobat files faithfully reproduce the printed page of the original print based article, and amount to little more than photocopies. Furthermore, most of these journals are print-based first and e-journals second. Submission of articles, editing and peer review are performed along

traditional lines, and the printed copy, along with its page numbers becomes the definitive "publication". What is envisioned for *Alternation* as part of the SALIT Web differs quite radically from this approach, placing the journal within a community of users who are actively invited to react, respond and discuss its contents with the authors and each other. Ideally, some of the responses would develop into full articles for later publication themselves.

An interesting case study is the journal *Glacial Geology and Geomorphology* (GGG) described by Whalley (Whalley et al. 1996: 171). The journal is what the authors call a "real" e-journal, i.e. there is no print-based version. Basing their decision to publish only electronically on the growing tendency of scientific journals to abandon traditional print publication, the authors also cite their own pragmatic reasons including cheaper distribution (especially where high-resolution colour images are required), and rapid publication. Amongst the questions asked by subscribers to the GGG was where the definitive ("Urtext") would be held and how it would be archived. The issue of definitive texts was a major preoccupation of the TEI project, which made mandatory the declaration of the source and revision history of each marked up text, as an answer to academics who were apprehensive about losing an indication of the provenance of a text, which in a printed book or journal is enshrined in the imprimatur. Electronic text (and its close cousin, hypertext) really is a very slippery medium compared to print - as I have discussed earlier, its very nature is antipathetic to the concept of fixed boundaries and closure. As such, electronic publication profoundly challenges closed institutions, including publishing houses, academic institutions and individual ownership of texts. Yet it is possible to harness the dynamism of electronic publication while retaining the conventions of academic publishing, as the editors of journals like the GGG have shown. If some of the perceived threats of introducing an e-journal are re-examined in the light of the many advantages, some very exciting possibilities emerge: for example, if peer review is regarded as the end of a process, rather than an absolute prerequisite for publication, a discussion group linked to the e-journal could be an

incubator for developing articles to be nurtured during their progression towards final acceptance. Quite aside from the value added by on-line discussion and advice, the networked environment contains hypertext linkages which will allow dynamic referencing between publications and other resources like the SALIT Web itself, thus giving users instant access to supplementary information. Perhaps the most significant difference between print-based and on-line publication is the considerable scope that the electronic environment provides for the inclusion of multi-media material. Music, the spoken word, video and animated diagrams lend themselves to computer-based publication, which may also offer linked data files, and translations into several different languages. Protecting intellectual property rights would to some extent be ensured by the moderated nature of the group, but there is no reason to expect that plagiarism would be any more prevalent in the electronic arena than it already is in print. Peer review for the GGG is still done off-line, by conventional post or by e-mail, but the editors anticipate using a "private" (password protected) Internet page for reviewers in the future.

Whalley admits that several assumptions have to be made about the reader of an e-journal, including computer literacy, access to a networked computer, and for those reluctant to read from a screen display, the availability of a printer. Part of the motivation for the SALIT Web project as a whole is the cheaper and more widespread distribution of South African literary texts, based on the current expansion of networked communication in South African universities and schools, and the plans for the extension of the Internet to the wider community, especially those in rural areas. At the beginning of the 1990s, networked distribution of material would have been seen as elitist but the spread of information technology has been so significant during this decade that it is the book - expensive, quickly outdated and environmentally exploitative - that is likely to be out of the reach of most people in the 21st century.

How could the SALIT Web combine an integrated discussion group and e-journal? Some answers to this question will already have become apparent in the foregoing discussion, but the various components of the system must be seen in relation to one another: the primary SALIT Web home page, the SALIT Web itself (containing the database, full texts, etc.),the SALIT discussion group and the *Alternation* e-journal.
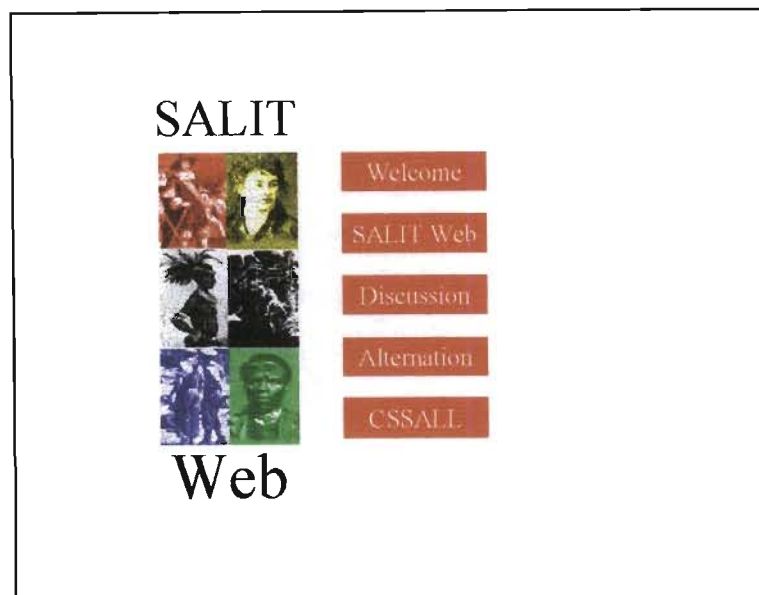
Figure 53: SALIT Web home page

The graphical interface in Figure 53 would lead the user to any one of the underlying access pages, but these, in turn would also contain further links across the various sites. We are concerned here with the combination of the discussion group and the *Alternation* journal, so I have confined the detail of the following diagram to those specific areas, bearing in mind that the other areas (in particular the SALIT Web) would have many cross links from their own hierarchical network structures. The "Welcome" page is provided primarily for the new user, and would give an overview of the site, including brief instructions to those wishing to

subscribe to the e-journal/discussion group and a description of the aims and scope of the group.

Further possible enhancements including a "closed" area for peer review and additional discussion groups to cater for specific subject areas have not been shown in the diagram.
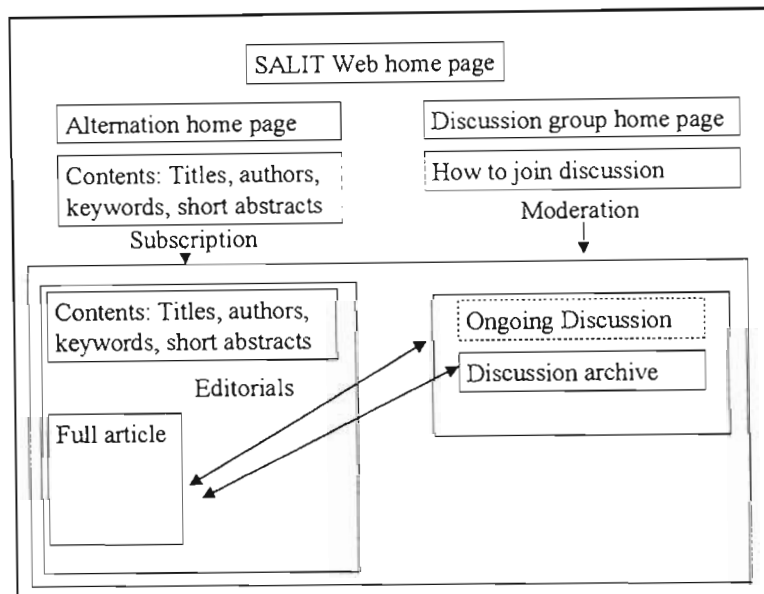
Figure 54: Integrated e-journal and discussion group

Both components are isolated from general Internet access by secure gateways that would require the prospective user to be registered by the site manager or the moderator. Casual visitors would be able to browse the contents and short abstracts of the articles in the various volumes of *Alternation*, but would be unable to proceed further before subscribing. The logistics of on-line subscription will have to be addressed by the project team as a whole, but the most likely solution would be one in which maximum value could be made available free or via a free registration, and only the full journals and discussion require a

subscription fee to approve members for the discussion group(s) and protect the intellectual property of the authors.

To test the structure and layout of an "electronic" *Alternation*, I used a disk copy of Volume 2.2 (1995) to reproduce the journal in a hypertext environment. Using Folio Views, I was able to convert the set of word-processed files into a Folio Infobase (textbase). As the disk-based version of the journal had been prepared for the printer with automated structural features in defined style sheet, elements such as titles, author's names, block quotations and footnotes were easily imported into the Folio environment. Whereas in my preparation of the Pringle text, every footnote had to be individually separated from the raw scanned text and marked as a footnote, the professional layout of the *Alternation* text allowed the footnotes to be hyperlinked to the body text as "pop-ups". Similarly, the article titles, already coded as first-level headings could be used relatively easily to generate a hyperlinked contents display.
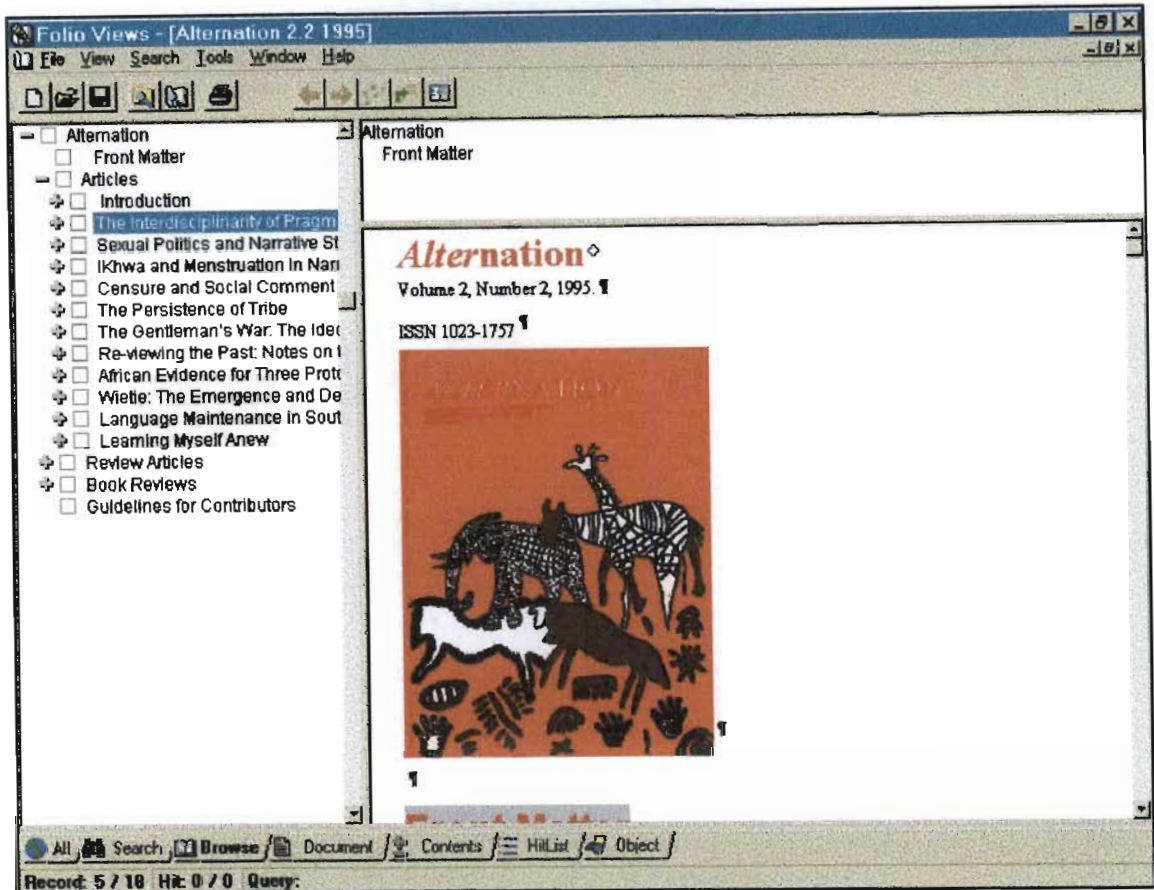
Figure 55: Opening screen of electronic *Alternation* journal

Selection of the article title takes the reader instantly to the full text, and clicking the display

tab allows full screen display of the text only. The powerful search ("query") features in the

Folio Views application described earlier in my discussion of Pringle's *African Sketches*

enable the reader to search a full word list of the text, or use the advanced query feature to

search for word combinations. In either case, clicking the "hit list" tab will produce a KWIC

display, showing the selected word or words in the context of the individual sentences in

which they appear. Were all the *Alternation* volumes to be made available within a single

web, it would be possible to perform search queries across all the titles, and find references

to, for example, "women" listed within the immediate context of the word, together with

the title of the article and the volume in which it appeared. A single click on the hit list entry takes the user to the point in the full text of the article where the word is used.

The prospect of locating the SALIT Web in the midst of a vibrant discussion group without geographical boundaries, where the participants may engage in debate sparked by publications in *Alternation* is an exhilarating and attainable goal. Apart from the value added by the debate itself, the published (and ongoing) discussions would be linked to the more formal structure of the SALIT Web database and virtual library, continually adding to the wealth of information in the resource as a whole.

In the next chapter, I document the development of the SALIT Web from a single chronology of South African literary titles in a word-processing file to a multi-faceted learning and research tool. The various stages in the design and implementation of the SALIT Web are outlined, and the solutions to some of the difficulties encountered in converting one format to the other are explained. The fairly tortuous process of importing linked information "chunks" such as date, title and author information into a new application was achieved without any loss of the integrity of the data, an important consideration for any other projects that may be undertaken in this area. Likely pitfalls and work-arounds are also detailed. However, the final result succeeds in bringing a virtual library of South African literature to the user's desktop.

Chapter 10 Development of the SALIT Web

Introduction

Because there is no established formula for the creation of such a resource, the design and development of the SALIT Web as documented here marks the end of a three year period of trial and error. The exploratory nature of the Web creation has generated several very different experimental versions since work started in 1996, each subsequent version a refinement of the last. The Beta 01 version, completed in September 1998, although fully functional (see Chapter 9) is really only one "snapshot" of a continually developing resource that has undergone quite substantial changes in size and sophistication (and name) both before and even since then. The actual process of development has been marked by recursive cycles of investigation, design, evaluation and implementation, followed by further investigation, re-design and so on. This chapter is an attempt to reflect on that process and to reveal the major lines of development, uncovering incremental improvements made over time. Just as in a hypertext web, this account marks out only one of many paths through a network of cross-references, doubling-backs and tangential excursions.

As a result, there is an apparent linearity to this account which belies the full reality. I have tried to record significant turning points in the development process so that others wishing to follow a similar route may use this as a guide, and to avoid pitfalls. Making the wrong decisions early in the project tend to produce very time-consuming, even disastrous, repercussions later. I have also painstakingly recorded every step at some of the more critical points because as the non-programmer soon discovers, computers demand very specific instructions, and software applications seldom deliver the seamless efficiency promised by vendors. Added to this, as an area of knowledge and innovation, electronic text in the humanities has advanced very rapidly over the period during which the SALIT Web

was constructed. My reading has revealed that there are still relatively few good examples of research and development in the area of humanities computing, but there is enough good practice to guide the novice. Significantly, now that electronic text projects in the United States and the European Union are starting to attract substantial funding, we have begun to see skills training in digital text production appear in undergraduate courses. To promote the sustainability of projects similar to SALIT, a draft curriculum for a course in electronic text production in our own institutions has been formulated and appears as an appendix to this thesis.

An attempt to map the broad pattern of development can be seen in **Figure 56**. I have attempted to regularise the terminology used in this account, but some apparent ambiguities need further explanation: in one line of development, the SALIT Web and the *South African*
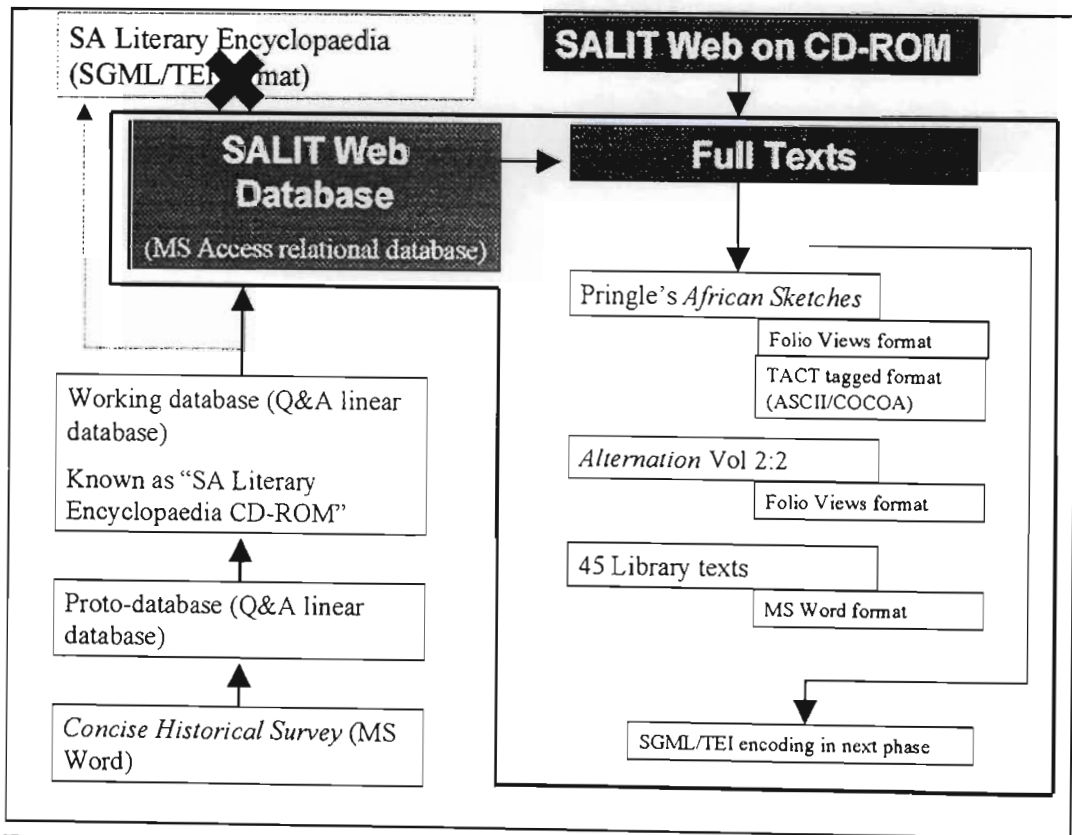
Figure 56: Overview of the development process

Literary Encyclopaedia CD-ROM were treated as one, hence at some points the database is referred to as an encyclopaedia (see Chapter 3 for further explanation of this issue). A similar name shift applies to the Beta 01 version of the SALIT Web, which is called "South African Literary Database" on the opening page. Most of these irregularities appear in the Figures which depict actual images of the project captured at specific times in its development.

Solving the Database vs SGML debate

Where to begin? At a fundamental level, electronic text restores the connection between the structure of writing and the practice of writing (and reading) that was undermined by the invention of the printing press. The printing and publishing industry delivers a fine finished product, but by removing the means of production from the writer, it has numbed our capacity to experiment with structures to express our thoughts. The industry has been a key part of the general social formation that treats books as commodities, and authors as the sole proprietors of the texts within them. While post-modern theory has challenged these notions, information technology has come up with hypertext, an ideal vehicle for subverting the authority of the printed word. On a superficial level, computers have already brought printing and publishing to our desktops, and so shattered the monolithic nature of the industry by allowing anyone to publish, if they have the equipment to do it. But the effect of hypertext is more fundamental, because it has opened up an unregulated medium of communication - the Internet - that cuts across the boundaries of printed text, organised publishing, the fixity of the book and the authority of the author.

Producing a hypertext requires the writer to work with the formal structure of the text as well as its capacity to radiate out into other texts. First and foremost, there must be close attention to structure: of the words in a sentence, of the sentence in the paragraph and up through other hierarchical levels to finally locate the full text in relation to other texts. Until

we have identified the vertical framework of the text explicitly within the dispassionate environment of cyberspace, it is impossible to conduct a meaningful exploration of the lateral relationships opened up by hypertext. In the light of this requirement - the systematic and explicit description of texts - the first noteworthy decision in the development of the SALIT Web was to use a database program to store the bibliographic information and commentary contained in the base document, Van Wyk's *Concise historical survey: South African Literature* (1996b). The intention of that publication was to give a diachronic account of South African literature. Electronic text was an ideal medium in which to explore the historical, linguistic and cultural changes reflected in the literature because it can provide multiple perspectives. By indexing each date, author, title and related commentary the electronic medium allows almost limitless re-orderings and groupings of the information. Databases are ideal for this purpose.

What is a database? An analogy is helpful here. In some respects, a database is like an ordinary office filing cabinet. Like a filing cabinet, a database can store information in ordered categories, allowing a user familiar with the indexing system easy retrieval of the information it contains. Unlike a filing cabinet however, a database can offer many indexing schemes at once, through electronic cross-referencing of its contents. To continue the analogy: a database is more like a multiple set of filing cabinets, all containing copies of the *same* information, but indexed in different ways: alphabetically, chronologically, by keyword, and so on. As in a filing cabinet, the database contents can be of two major types: either actual files of information containing pages of text, photographs and notes, or simply note-cards (like bibliographic card catalogues in libraries) containing only summary information that direct the user to textual or graphic material elsewhere (for example on library shelves). In a database, information can be collected and organised just as in a card system, but it can also permit sorting and selective retrieval of the information with great speed and accuracy. As an indication of the organisational capacity of a database, for the bibliographic information in the SALIT Web alone, there would need to be at least twenty-

five separate filing cabinets, arranged in alphabetical order by author's name, date of publication, etc. to accommodate the current range of attributes (or "fields"). To retrieve a simple combination of attributes - published work by women authors within a given time period, for example - a manual method would require all the records to be re-sorted, and would rely on the visual acuity of the sorter to stumble across each record where the gender was female and the publication date fell within the required boundaries. Retrieving records according to multiple criteria can be achieved faster and more accurately on a database. Before exhausting the filing cabinet analogy, we might consider the multiplicity of duplicate records necessary for the other major component of the SALIT Web, the full text library. To produce a KWIC display of the occurrence of the word "river" in Pringle's *Narrative of a Residence in South Africa* (1834) one would have to imagine the logistics of visually scanning every page and then creating a new set of records sorted according to each page or paragraph on which the word "river" appeared. It would probably take several weeks to produce all the information contained in the word frequency lists that can be seen in Chapter 7. With their ease and speed of retrieval, database structures offer an ideal medium for the reiterative exploration required in the SALIT Web.

Database specialists with literary-historical backgrounds are rare. One of them is the historian Daniel Greenstein, (Greenstein 1996b) whose lucid description of structured databases applied to humanities studies had a crucial impact on the development of the SALIT Web - the decision to retain a database structure for the bibliographic portion of the SALIT Web, while continuing to aim at SGML/TEI encoding for the full texts in the Library section. The flexibility and comprehensiveness of the database is a key element of the completed CD-ROM.

Starting with an account of database types from simple (linear) databases through multi-table (relational) databases to process-independent designs (e.g. the TEI), Greenstein considers the relative strengths and weaknesses of each and analyses their research

potential. He stresses the importance of defining data types in terms of the research question or assumption, *before* determining fields, as well as the consistency of their application once data is being added. The limitations of the single-table databases - like the Q&A Database, the "working" database that we used in the very first stage of the project - include their inability to handle observations or information of varied richness, and the absence of mechanisms for recording ambiguous material. Another drawback of the single-table or "flat" database is the difficulty posed in adding new fields and the related problem of irreversible data coding and structuring decisions, an important consideration for the SALIT Web which was intended to be a dynamic, growing resource that could be adapted over time. The advantages of the relational database, influenced the decision to abandon the idea of an SGML encoded structure for the bibliographic level of the SALIT Web.

Unlike "flat" databases, relational databases are multi-tabular in design and so allow the creation of separate tables containing records pertaining to a particular "object", thus making it easier to manage sequential attributes, i.e. objects which have several attributes become new tables, while the relationship between records in two or more tables is determined by the values stored in fields they have in common. The implications of this multi-tabular capability clearly had profound implications for the design of the primary SALIT database, demanding as it does a one-to-many relationship between, for example, author and title entries. Even more significant is that the complexity offered by a relational database is comparable to an SGML encoded document. Whereas the inadequacy of the single-table Q&A database used for the original SALIT content meant that it could be viewed only as a temporary measure leading to the later development of an SGML encoded text, the considerable versatility inherent in a relational database suggested that a database could become a permanent feature. It soon became clear that the Q&A version of SALIT would have to be converted into a relational one.

Greenstein had indicated that a "process independent" design (like SGML), although more complex than a multi-table database might not be appropriate for all types of information, especially that which was already contained in database form. There would seem to be no point in converting a database to SGML when the material contained in it was already structured and encoded (the example of Peter Robinson's conversion of the *The World Shakespeare Bibliography. 1990-1993* notwithstanding. See page 204 below). Such a decision would not, of course, accommodate another aspect of the SALIT project plan which called for the incorporation of full texts into the CD-ROM version.

Adopting separate structures for the bibliographic and full text components of the SALIT Web resolved the ambivalence we had experienced in the first stages of the project over the relationship between the SALIT Web on the one hand, as a form of enhanced bibliography, and on the other, as a virtual library of full texts (primary texts, critical commentaries and reviews). SGML was so clearly the established and appropriate method for encoding the texts, that the same appeared to be true for the SALIT Web as a whole. After all, distinguished precedents already existed for using SGML for this type of document: J.L. Harner's *The World Shakespeare Bibliography. 1990-1993* CD-ROM and the *Oxford English Dictionary*. Greenstein's article, and one by Peter Robinson that discussing *The World Shakespeare Bibliography. 1990-1993* CD-ROM (Robinson 1996), clarified the apparent ambiguity in the initial design strategy for the SALIT Web. I was later able to verify this decision with Peter Robinson himself when I had the opportunity to discuss the SALIT project with him at a conference. Greenstein's database paper together with Peter Robinson's account of the development of *The World Shakespeare Bibliography* have, therefore, been crucial to developmental decisions taken subsequently in the SALIT Web project.

An account of the process

From the print-based Survey to the working database

The following discussion traces the (often uneven) process of converting the *Concise historical survey: South African Literature* (Van Wyk 1996b) from print-based to electronic database form, and then the creation of the SALIT Web as an integrated whole.

Turning the word-processor file containing the *Survey* was an exacting operation that had many false starts. The first consideration was finding a way to preserve several print-based elements in the original by representing them in machine-readable form. Amongst the features of the original that demanded translation were font indicators of genre and canonical selection; specific author, title, date and biographical information; and references to events. In database jargon, some of these would have to be identified as entities (or "records") and others as attributes (or "fields"). Andrews and Greenhalgh (1987) have claimed that if "entities" is a noun, then "attributes" is an adjective. I use the terms interchangeably in the following discussion.

As has been mentioned, at the start of the project the creation of the database version was viewed as an interim measure to structure the SALIT material prior to encoding it as a full SGML "mega" document. There were two reasons for this decision, firstly the delay in obtaining suitable SGML editing software and browsers at the beginning of the project and secondly, the urgency of continuing with the capture of entry data. Because the SALIT Web was at first closely linked to the development of the print-based *Encyclopaedia,* a comprehensive list of title and author entries was needed to assign writing tasks to potential contributors. Having secured the endorsement of a range of academics and publishers at a preliminary workshop in June 1996, the editors were in a position to proceed with the identification of project leaders for the various encyclopaedia sections, and to allocate

entries to a large number of individual contributors. Hence, the contingencies of managing the larger project (the *South African Literary Encyclopaedia*, per se) eclipsed the immediate requirements of the associated electronic text (CD-ROM) version project. However, the momentum of the main project proved eventually to be of benefit to the CD-ROM endeavour as it served to accelerate the growth of the database.

A brief consideration of the distinction between the SALIT Web as a *database* and as *SGML* electronic text is needed at this point to justify the decision to expend considerable time and effort in producing what are in some respects duplicate electronic versions. To begin with, in electronic text terms, a database is old technology while SGML encoded text is new. But there is more to this distinction than mere novelty. While databases are designed as highly versatile "index card" systems, they are essentially useful only where information can be structured as a collection of records sharing the same well-defined fields. Attempts to represent "running prose" expressing impressionistic data or rich descriptions of individuals, books or events reveal the limitations of a database, especially the linear single-table type such as Q&A, which was used at the beginning of the project. SGML on the other hand, which also encodes into entities and attributes, is directed specifically to a description of the internal structure of texts. SGML is "process independent", allowing the indexing and cross-referencing of textual structures and units without having to conform to the rigorous and relatively inflexible requirements of database software. Once the structure of the text has been encoded, the user can continue to tag successive attributes of the text, the one nested inside the other, right down to the morphological level. Then, by using a text analysis program such as TACT, researchers have the tools to generate their own indices, word-counts, collocations and so on. While the database assigns key words *en bloc* to entries, SGML allows various levels of enquiry within full texts, that may be customised to the needs of the individual researcher. While a database forces rigid categorisation, with all the limitations this implies (the list of keywords is limited, and linking them to already captured entries entails an unfeasably long re-encoding of existing data), an SGML encoded

text lends itself dynamically to new keyword searches - the nature of the encoding allows indexing, search and retrieval software to make associations automatically and present them for immediate use by the researcher. Given the flexibility inherent in SGML markup, it seemed at this point in the development of the project that the database should be regarded as just a means to an end, and that the entire SALIT Web would eventually be produced as a very large SGML "mega" document. As has been noted, this decision was to be reversed once the SALIT web had been re-conceived as a two-tier entity containing a bibliographic database *and* a virtual library of full texts housed within one hypertext environment. Although the intention to create an SGML "mega text" was to some degree a blind alley, it proved to be of considerable assistance in conceptualising different types of representation, essential to the conversion of the *Survey* to the database, as I will show later (see page 188 below).

Before the decision was made to retain the database in the SALIT Web, the advantages of using a database soon became apparent in the project as a whole. As the project management features of the working database became clearer, it could be seen that in addition to the bibliographic information, it could also store the author biographies submitted by contributors to the *Encyclopaedia*, and even the names and addresses of contributors. The database could also be used for mundane administrative tasks like printing address labels, etc. Thus, the database version, although regarded at first as something of a tangential move in the overall development of the electronic version of the SALIT Web promised to be a support system that would benefit the overall project in several useful ways, not the least of which was its potential for making the bibliographic information (and the author biographies) available on-line via the Internet to *Encyclopaedia* contributors.

The methodology for the conversion of the existing version of the encyclopaedia had to be determined according to the contingencies of the situation, and some unique solutions were found for tagging and ordering the attributes so that they could be downloaded into the

database. Johan van Wyk's original document *Concise Historical Survey: South African Literature* (1996b) which was to form the basis of the new encyclopaedia and the database, had been produced as a MS-Word document, complete with font formatting that indicated canonical works, poetry, prose and drama. Here, the intention to represent the final version in SGML was to prove invaluable. The first step in the conversion was the encoding of the *Survey* document with broad TEI element tags, which involved the insertion of tags throughout the text of the following SGML conventions: <persName> (used for all authors and other references to names), <title> (all publication titles), and <date>. The following tags were also introduced to further refine the coding: <event>, <prose>, <poetry> and <drama>.

While this encoding continued, both within the text of the original document and then for additional entries, the task of transferring the text into a form suitable for downloading in to a database was undertaken.

The first step was to make a copy of the original MS-Word document, dated to pinpoint the parameters of the database content. It was decided that work on adding bibliographic information to the "original" MS-Word version of the *Survey* would proceed uninterrupted, thus placing the conversion part of the process under severe time pressure. While the "original" was being converted, a new version was simultaneously being added to, necessitating an additional conversion later. In retrospect, it would have been wiser to have suspended the data entry until the database was ready, as this parallel work complicated the transfer of information and could have led to the omission or duplication of data.

After the MS-Word copy had been made, it was converted to ASCII to eliminate all font and formatting enhancements. As mentioned above, these originally had an "encoding" function themselves, but this had been superseded by the SGML tags, which by their nature are software-independent and capable of surviving conversion from one word-processing

application to another (and, of course into ASCII). As already noted, without the initial SGML tagging, the conversion of the text into database form would have been well-nigh impossible, underscoring the value of such tagging even when it is not used for directly SGML purposes. The value of the tags lay in their consistent signalling of categories of data in the ASCII text, allowing the use of a word-processor search-and-replace function to find each occurrence and enter a "field"-like indicator - in this case a paragraph-end code.

A parenthetical note for anyone intending to use a similar method to convert running text to database form is that the choice of word-processing programs is an important consideration. An unavoidable complication at this point of the process was my relative unfamiliarity with MS-Word, which necessitated the actual searching and replacing process to be carried out in WordPerfect. An early trial run proved that the database (Q&A) did not recognise the version of ASCII saved in WordPerfect which in turn required that the document again be converted into MS-Word to prepare it for export to the database records and fields. However, the meticulous naming and tracking of files ensured that the process could be performed systematically, while back-up versions could be stored in case of certain steps having to be repeated.

To continue with the account of the conversion: once the paragraph ends had been inserted, the text could be "cleaned up" by the removal of information unnecessary for the database conversion. A balance had to be found between the ideal of transferring everything in the original to appropriate database fields, and the time constraints that applied. Further, the information deleted from the document would not be lost, but could be re-inserted into the database once its basic structure had been created, and the primary content transferred in a suitably ordered way. A key decision at this stage was the determination of which of the tagged fields would constitute the primary entity (record). The most feasible route was to treat each existing *Survey* entry as the basis of each record, and although these were presented chronologically in the MS-Word document, the predominant entity was the Title.

So it was determined that each record of the proto-database (including Events) would contain four fields, namely, "persName", "Title" and "Keyword", with the first field to be extended to "Surname" and "First Name/s" prior to the actual download. Because the decision to use the broad TEI tag <persName> (without embedding <forename> and <surname>) had been made before the decision to create the database, the tagging was not sufficiently refined to fully automate the process. The consequences of this omission are referred to below.

The "cleaning up" of the unnecessary text was the most time-consuming part of the process as the "tags to paragraph end" conversion had not taken account of the text following the title, thus leaving approximately half the entries with "title" fields that contained genre-like indicators, e.g. "short stories" or "anthology" and biographical commentary, e.g. "this work was awarded the Hertzog prize in 1965". Another difficulty was the large number of entries (approximately one third) that reflected two or more authors. The "rough sieve" of the search-and-replace had included all of these in one undifferentiated field "persName".

The solution devised for the "persName" inconsistencies allowed at least one surname to be isolated for the "surname" field and the rest of the entry to be included in the "First Name/s" field: an awkward solution, but sufficient for a basic database to be created, and within the database structure, flexible enough to allow for accurate retrieval of second and third author information. A macro was written that searched for each "persName" occurrence, highlighted the last word of the entry and transferred it to a new field tagged as "surname". Although successful for most entries, exceptions were two- and three- word surnames such as "du Plessis" and "van der Merwe" and in two- or three-author entries, (in for example, Adey, Beeton, Chapman and Pereira 1986) where "Plessis" or "Pereira" would inaccurately appear in the primary surname field. In the final download of the database information, these inaccuracies had to be manually cut and pasted, for almost one thousand entries.

By far the most time-consuming part of the process was regularising the "title" fields. Many entries in the 16th, 17th and 18th Centuries had extensive commentaries, most of which were essential to a full appreciation of the title entry. The commentaries are a unique feature of the *Concise Historical Survey* and could not be eliminated without compromising the essential nature of the text. As there was no way of treating this data automatically, each entry was examined to determine whether the information it contained had been repeated elsewhere, and then either retained in full, edited or deleted. This decision was influenced by the interim nature of the proto-database as none of the material would actually be lost in the final version of the database. As a result, many of the first 4009 entries had additional material included in the "title" field that would at a later stage be transferred to the "keyword" and "commentary" fields.

Another decision that had to be made on an item by item basis, was the inclusion or deletion of the entries originally tagged as <event>. Like the commentaries on individual titles, these entries are part of the unique structure of the *Concise Historical Survey*. As the design of the proto-database excluded an "event" field for the time-constraint reasons given above, where were these to go? A compromise was the inclusion of most of these "events" in the "title" field, even though most of them were not publications at all. The decision to exclude an "event" field in the proto-database proved to be a poor one. Two major difficulties arose as a result of this decision: firstly, the 4009 entries obviously contained some "titles" which were actually events, and secondly, trial runs of the export to the database revealed that empty fields caused inconsistencies in the download, scrambling the fields in random ways. To overcome the latter problem, it was necessary to add "title" field content to the event fields, compounding the problem mentioned earlier, by including event information in some of the "Surname" and "First Name/s" fields as well. Although this does not prevent successful retrievals, it violates the principle of unique field identities that is one of the major strengths of a database. Once the working version of the database was in place, a separate "event" field was added to regularise this inconsistency.

In retrospect, the omission of the "event" field in the proto-database to speed up the transfer process proved unsuccessful. Up until the final download, the very last step in the process, analyses of the data showed that although there were 4009 entries (dates, names and keyword fields) there were only 4007 "title" fields. Search functions in both MS-Word and WordPerfect proved inadequate to explain this anomaly. Three physical scroll-down searches of every entry also failed to pinpoint the problem (but these laborious searches did uncover another hitherto undetected problem associated with paragraph-end codes, see the next paragraph). It took approximately ten working hours to perform these manual searches, far longer than it would have taken to have isolated and encoded an "event" field.

The undetected error which showed up in the manual searches had affected the reliability of the trial database exports earlier. The transfer from word-processor to ASCII and back had converted soft-return line-ends into hard paragraph tags in longer fields. This had the effect of indicating during the database export process that there were more than four fields, thus scrambling the results. A partial solution to this problem was using the "text only" rather than the "text with line breaks" save option in MS-Word, but the frequent conversions from one word processor to the other resulted in inconsistencies in the retention of the soft-returns in the ready-for-export ASCII text. The inconsistencies were eliminated in two steps. The first step was the re-formatting of the document into a landscape page layout with minimum left and right margins, together with the smallest font size. This effectively lengthened each line, preventing word-wrapping and the inclusion of soft returns. The text could then be visually scanned for inconsistencies and incorrect paragraph-end codes deleted. Later discussion of this difficulty with Wendell Piez of the Electronic Text Centre at Rutgers University suggested that a possible solution to this problem would have been the use of a powerful ASCII-based word processor (e.g. Xywrite) from the outset. Even with the page widening, further visual searches performed to find the elusive missing "title" fields mentioned earlier revealed further unwanted paragraph-end codes. Proofing of the

working version of the database revealed further instances of this anomaly that had escaped the screening process.

The final stage in the conversion of the text to the database was the export process itself. Trial runs showed that either the memory capacity of the computers available or a limitation in the Q&A database prevented more than 783 entries being downloaded in one run. As the reliability of the process was being compromised by attempting to export all the records at once, it was decided to split the document into eight 500-entry files, making each export procedure fit comfortably within the proven tolerance levels. Once the files had been split, details of the start and end entries had to be recorded by hand so that each export could be verified. Because the ASCII text had by this time been largely regularised by the elimination of inconsistencies and the systematic presentation of fields, the process was successful and relatively quick. It was only at this stage however that the missing "title" fields were detected and corrected. Analysis of files 7 and 8 showed each to be one "title" field short. By using search and replace to tag each title field with a conspicuous symbol, a visual search of the records soon revealed the missing field in file 7, which was found to be incorrectly tagged as "event". A search for an <event> tag in file 8 found the second inconsistency.

The SGML tags remained with the text until just prior to the export, as they continued to serve a useful purpose in giving the document clear field identification tags. The end-tags (e.g. </title>, etc.) had been deleted at the start. The only tag fragment to be exported was the word "key" (without angle brackets) which ensured that none of the keyword fields, irrespective of any keyword having been entered, appeared as a blank field (see comment on empty fields, above).

The "keyword" field was an addition to the original text, which ensured the immediate use of the working database to cross-reference form, genre, language characteristics, amongst

others. Although it was not feasible to enter a comprehensive list of keywords during the MS Word-to-database conversion there was some opportunity for adding keywords at various stages during the conversion and export process described above. The value of the keyword feature in the SALIT Web is discussed in Chapter 2. The following table shows the keywords used during the conversion process (these were subsequently augmented for the working database, and now there are just under 5000 separate keywords embedded in the database records as well as over fifty "topic" searches based on multiple keyword combinations):

afrikaans	history	public speech
autobiography	journal	san
award	legislation	satire
ballet	letters	sechuana
bibliography	music	sotho
biography	newspaper	translation
conference	novel	travel
diary	oral	voortrekker
dictionary	organisation	women
drama	pamphlet	xhosa
event	poem	zulu
herero	prose	

Figure 57: Keywords in the first working database

Certain keywords were entered consistently so that the retrieve function of the database could be demonstrated as soon as the working database went into use: all travel, Sotho, translations, legislation, journal, newspaper, dictionary, bibliography, biography and autobiography entries were entered. Less consistently entered were: poetry, drama, Zulu, Afrikaans, Xhosa, letters, diaries and history. The balance of the keywords for the original 4009 entries were to be entered once the working database was in use, while all new entries would have keywords entered from the augmented list. Apart from their utility in the database, the keywords contribute to other structures in the SALIT Web such as the

learning applications, and will be incorporated into the TEI encoding of the full texts in future stages of the project.

The working database

The proto-database had four fields as described above: Surname, First Name/s, Title and Keyword. These were incorporated into an input form, designed as a front-end to the database for the input of new data for retrieval purposes. Also on this page, signalling our preoccupation at that stage with the "intermediate" status of the database is the heading "South African Literary Encyclopaedia CD-ROM (non-SGML database version)". Although in retrospect our insistence on this distinction seems rather pedantic, it had the effect of reminding the project team that in spite of its advantages as an information retrieval device, the working database was quite different in structure from the intended final electronic version (for the reasons outlined above). Until the later separation of a relational database as a "finding aid" and the SGML-ready "full texts" now reflected in the SALIT Web, it was still felt that if the *Encyclopaedia* CD-ROM - as it was then called - and its associated full-text contents were to conform to international standards (and so allow future sharing of resources, and inter-database searching) it would have to be encoded as one large TEI/SGML document.

Three more input forms were added to the "General Input" (Author and Title) form shown

Figure 58: Front-end of the original working database (no longer used)

in **Figure 58** : Reviews, Full text and Events. Although the retrieve facilities of Q&A were sufficiently flexible to find such information from any of the fields, it was decided that the usability of the database would be enhanced by presenting this selection of front-end possibilities organised in a semi-hierarchical structure from the Author/Title "main page" through to "sub-pages" devoted to the other three categories.

From a linear to a relational database

Amongst the reasons already mentioned for migrating from the linear (flat) Q&A database to a relational database, is its multi-table structure. The working database was essentially only one table, each entity (record) reflecting one "object" - the title of a book or an article, with additional attributes (fields) providing the author's name, gender, etc. While this was adequate for the systematic addition of bibliographic information, it provided little scope for the inclusion of multiple levels for any one attribute: how would one include the dates for revisions or new editions of the same title? There was already a formidable amount of repetition in the database - author name, for example, was repeated in every title entity. To accommodate dates for revisions, new attributes would need to be added to the table, but how many? Some titles might require a large number of distinct attributes for first, second, and third revisions and so on, while others would require none. The solution, as indicated earlier in this chapter was the adoption of a relational database.

By converting the linear database to a relational one, several tables could be created, each linked to the other allowing a table exclusively for author entries, another for titles, another for events and so on. The conversion would allow for much greater flexibility, and because the title table, for example, would now be separate from the author table, it could accommodate any number of new titles and revisions joined to the "parent" table by values stored in common fields.

ID	surna	first	date	book title	publisher	ge	key
7673	Pringle	Thomas	1828	Ephemerides; or, Occasional Poems,	London: Smith, Elder	m	poem;

Figure 59: Entity (record) in Q&A linear database (truncated)

In the single-table entity from the original Q&A database (Figure 59), all the attributes are clustered together, while in the new multiple table database (Figure 60) the arrows indicate how the two entities are joined by the values stored in the common Author_ID fields.

ID	Author1_I	surname	first name	gender
72	6725	Pringle	Thomas	m

ID	Author1_ID	surname	first name	date	book title
7269	2764	Greig	George	1827	The colonist
7672	6491	Philip	Rev. John	1828	Researches in South Africa.
7673	6725	Pringle	Thomas	1828	Ephemerides; or, Occasional Poems, written in Scotland
65	7836	Stapleton	R.J.	1828	Poetry of the Cape of Good Hope.
67	8090	Swaving	Justus Gerhardu	1829	Vlugtige biografische schets
66	7034	Rose	Cowper	1829	Four years in Southern Africa.

Figure 60: Two joined tables from the new relational database (MS Access)

Owing to the already systematic structure of the old working database, the task of importing the data into the new program was relatively uncomplicated. Even though the new MS Access database could not "recognise" the Q&A file format, the old database had a facility for converting its files to generic ASCII that could in turn be imported into MS Access. Once the conversion had been completed, the new database contained a copy of the single undifferentiated table that could then be split into two or more related tables. Using a copy of the composite table and by assigning a unique ID to each author and another to

each title, I was able to use the "delete duplicate records" function in MS Access to eliminate multiple occurrences of each author entry. All other fields were then removed, and the tables joined by the values in the common Author_ID fields. Figure 61 shows the imported table (qabase_OLD), the interim duplicates table (qaduplicates). Also shown are the new tables (Author and Title) and more recent additions created after the initial conversion. The thin arrows indicate the joined fields.

Figure 61: Tables and relationships in the new MS Access database
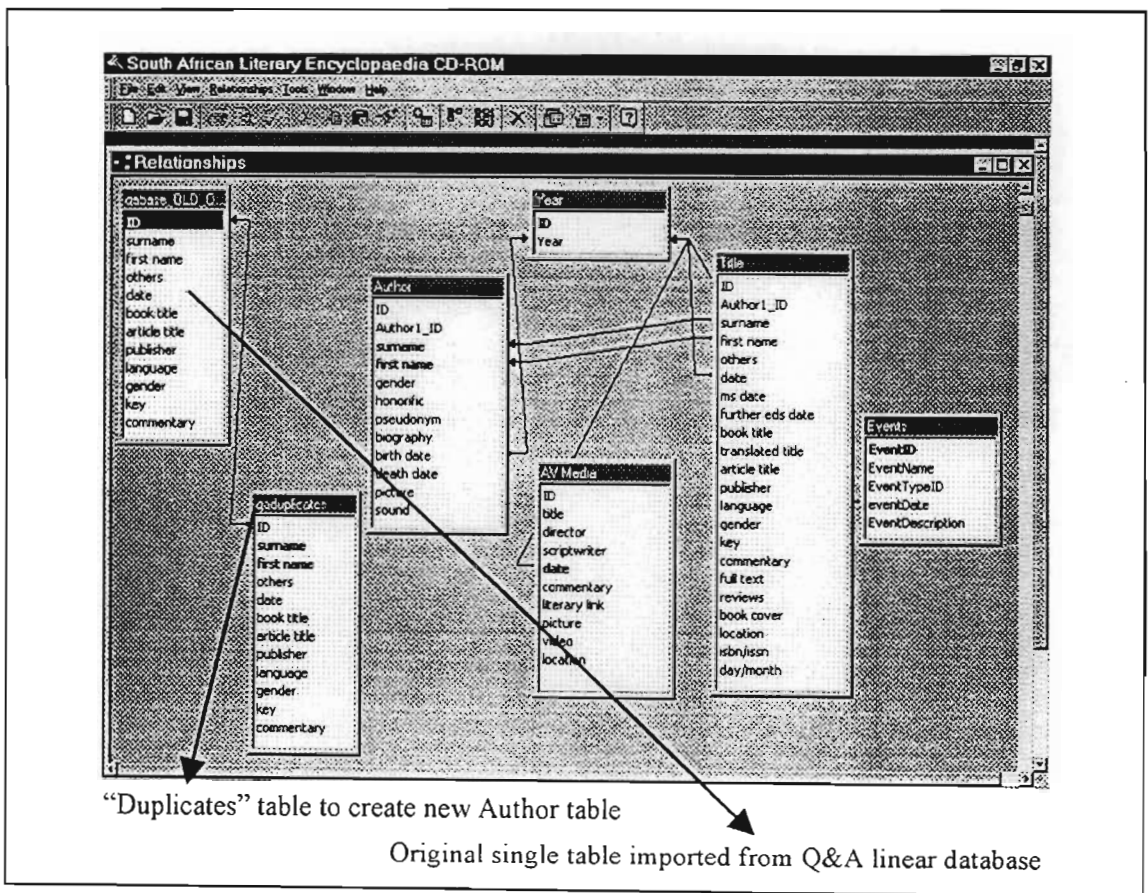
A step-by-step account of the above process is included as Appendix B.

Towards Internet access to the database

It was determined by the leaders of the *Encyclopaedia* project that the Centre for Southern Literature and Languages (CSSALL) could greatly enhance communication with remote contributors to the Encyclopaedia by having an Internet site that would contain a description of the project, list the names of contributors, supply e-mail addresses, and enable on-line searching of the database as it developed.

Thus the next phase in the development of the database part of the SALIT Web was the design, data collection and encoding of several documents that would constitute a CSSALL "home page" for publication on the Internet. Information relating to the SALIT Web and *Encyclopaedia* projects would form a number of subsidiary web pages accessible via the main home page. Only the SALIT Web and *Encyclopaedia* pages are discussed here.

The first publication of these pages took the following pattern: a non-linear web of pages, accessible by hyperlinked anchors from the CSSALL main page under the headings "Research Projects" and "South African Literary Encyclopaedia". The *Encyclopaedia* page was further linked to three other pages: an overview of the project (consisting of an early draft of Chapter 1 of this thesis), a page providing guidelines for contributions to the Encyclopaedia (word-processing format, file name protocol, etc.), and a further page providing a list of Internet links to South African and international sites containing information on electronic texts and projects.

Encoding of the pages was in standard HTML, suitable for Internet browsers such as Netscape or Mosaic with minimal use of graphics to ensure quick download times. Prototypes of the pages were viewed locally and amendments and adjustments made before arranging to formally "publish" them by mounting them on the University server. Further extensive editing of the pages was necessary during the "publishing" stage of the procedure,

as it was decided to standardise file name suffixes (as "*.html"), and change the main page file name to "index.html". This required re-writing all the internal anchor hypertext links and testing them. The current location of the files is on the University of Durban-Westville "Pixie" server in /home/stewartg/public.html. The pages were published on 28 August 1996, and followed up with an e-mail message to all contributors informing them of the Internet address: http://www.udw.ac.za:80/ ~stewartg/index.html.
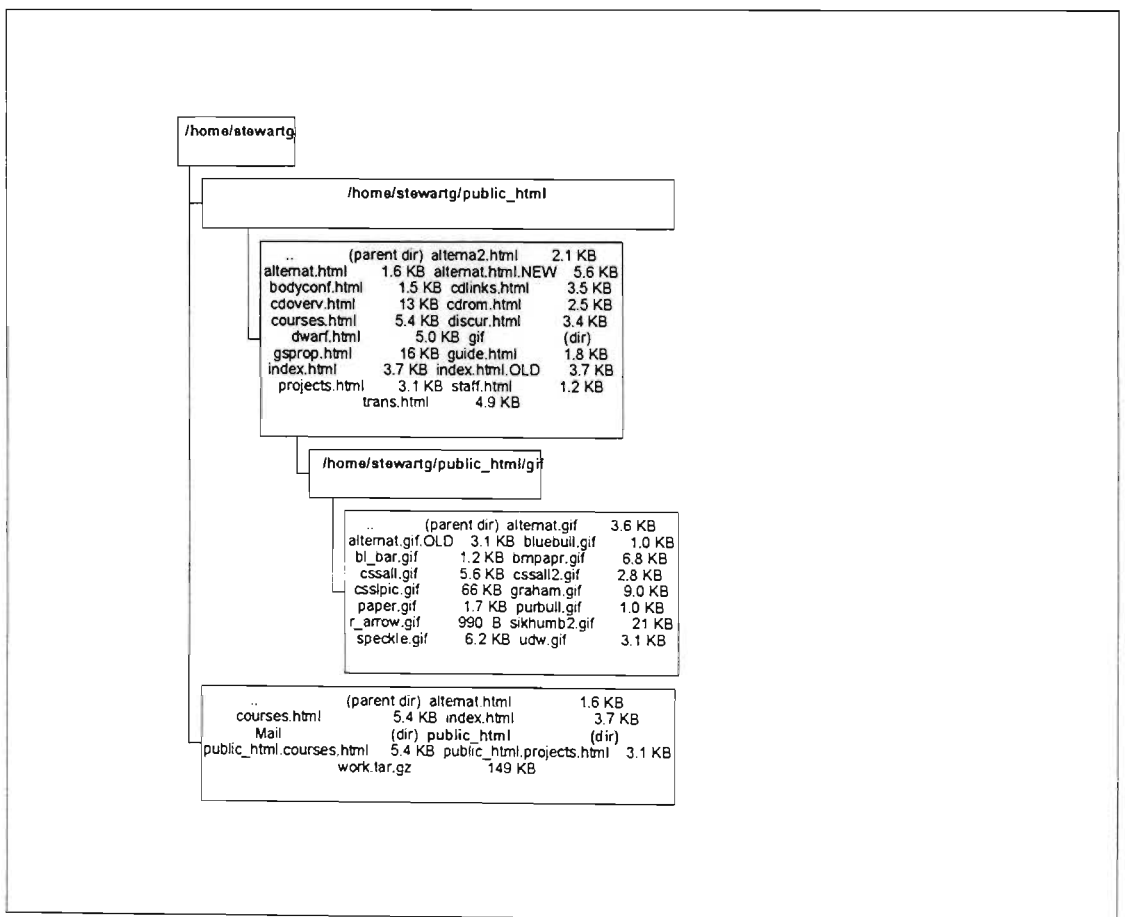
Figure 62: CSSALL Internet site directory structure (on the university server)

Although the next stage of the home page project was to be the development of a Common Gateway Interface (CGI) allowing contributors to access the database, we were unsuccessful for reasons that follow. The Internet access had eventually to be postponed

while funds were sought to establish a server locally with the CSSALL. The original plan was to begin with a design that limited queries to only three fields: author, title and date. This modest interface was justified by its initial intended use: as a "sampler" of what the SALIT Web CD-ROM would eventually be able to provide, and as an interactive information source for contributors. It was hoped that with the publication of the first CD-ROM, the interface could be extended to provide accessibility to information similar to the CD-ROM itself.

Figure 63: Internet gateway for the planned CGI access to the database (not implemented)

The development of the CGI was undertaken in collaboration with Simon Tyrrell, the Network Manager at the university, who was able to bring his programming knowledge to

bear on the design. To begin with, an HTML file was created from the new MS Access database. The intention was to store this "snapshot" of the database on the university server, and have an existing Grep programme perform the search. However, the resulting HTML file proved to be unmanageably large, which would have resulted in unacceptably lengthy internet searches. A solution appeared to lie in the design of a "flat" text file, similar to those that had been used earlier in the project for the interchange of database information from the *Concise Historical Survey* document to the single-table Q&A database, and later to the relational MS Access database. To save space, and for the reasons mentioned above, only author, title and date fields were to be included in the text file. Despite the reduction in the file size, the database contents proved too big for the Grep program, and Internet access was postponed until a server could be established within the CSSALL. Work on this part of the project is due to re-commence as a server was installed in the Centre at the end of 1998.

From database to SGML/TEI: a false start

From the discussion of the two-tier database/full text structure earlier in this chapter, it will be clear that the SALIT Web no longer aspires to unitary SGML structure. Yet for at least the first year of the project, the publication of the SALIT Web encoded in SGML/TEI was the central aim of the project. Some progress was therefore made towards the realisation of this aim, starting with the primitive SGML markup during the transfer of the *Concise historical survey: South African Literature* (Van Wyk 1996b) to database form (see above). The combined effect of the contingencies of work on the database together with inadequate skills and a lack of the appropriate software, delayed further progress towards full SMGL/TEI encoding. However, my attendance of the CETH (Center for Electronic Texts in the Humanities) Summer Seminar at Princeton University in July 1996 provided the latter resources, and a pilot SGML version of the SALIT Web (in the form of an electronic encyclopaedia) was undertaken at Princeton under the guidance of Wendell Piez of Rutgers University and with course material prepared by Michael Sperberg-McQueen, co-director of

the TEI. Further guidance was obtained from Peter Robinson, editor of *inter alia*, Samuel Johnson's *A Dictionary of the English Language* CD-ROM who had advised on the adaptation to electronic form of *The World Shakespeare Bibliography* (Harner 1996). Discussions with Robinson, and an examination of the original form of the *Shakespeare Bibliography* information, pointed to the adoption of a similar TEI DTD (document type description) for the SALIT Web. As had been the case with our own "interim" database version of the SALIT Web, the text of Harner's Bibliography had been stored in a database which the staff at Texas A&M University had then subsequently used for input into the SGML version.

This is an example of one such database record:

```
%20.15
%.10 Sahel, Pierre
%.25 'Monarques et mendiants shakespeariens.'}
%.30 {Difference<61>et identite<11>} [F]: 105-19.
%.35 1992
%.40 [Using the English histories, {iKing Lear}, and {iTimon of Athens}, explores how
Shakespeare represents the identification of king with beggar.]
% .64 beggar, monarch
% .80 9315artsh
```

(Robinson 1996: 4)

Compare this with an example record from the SALIT database:

Field 1(date): 1819

Field 2 (surname): Pringle

Field 3 (firstname/s): Thomas

Field 4 (title): Autumnal Excursion and Other Poems

Field 5 (commentary):

Field 6 (keyword): poem; colonial;

Field 7 (full text): The Forester of the Neutral Ground ...

Field 8 (picture): pringle.bmp

The basic structure is similar, although Harner's *Bibliography* had a primary division into "General Shakespeareana", "Play Groups", "Individual Works" and "Indexes", with the "Play Groups" further sub-divided into "Histories", "Tragedies", etc as in the example. In converting the *Bibliography* information into SGML/TEI form, Robinson decided that the TEI <BIBL ID> tag was flexible enough to be extended from its usual function of tagging citations in the text to providing all the bibliographic information required by Harner's project. The encoding of the example record given above into TEI/SGML is as follows:

```
<DIV ID>='w94-Pla' CODE='w94-20'>
        <HEAD>Play Groups</HEAD>
        <DIV ID='w94-Pla-His' CODE='w94-20-15">
                <HEAD>Histories</HEAD>
                <BIBL ID='w942015SahPieMonet' N='143">
                        <RESPSTMT>
                                <NAME TYPE='principal'> Sahel, Pierre.</NAME>
                        </RESPSTMT>
                        <TITLE TYPE='part'> 'Monarques et mendiants
                        shakespeariens.'</TITLE>
                        <IMPRINT><HI REND='ital'>Diff&eacute;rence et
                        identit&eacute;</HI> [F]: 105-19.
                        </IMPRINT>
                        <NOTE TYPE='ed'> [Using the English histories, <HI
                        REND='ital'>King Lear</HI>, and <HI REND='ital'>Timon
                        of      Athens</HI>, explores how Shakespeare represents
                                the identification of king with beggar.]
                        </NOTE>
                        <KEYWORDS>beggar, monarch</KEYWORDS>
        </BIBL>                                          (Robinson 1996: 5)
```

The pilot structure for the Encyclopaedia developed at the CETH seminar using an adaptation of the above model was as follows:

```
<!DOCTYPE TEI.2 PUBLIC "-//TEI//DTD TEI Lite 1.0//EN" "TEILITE.DTD">
<TEI.2><TEIHEADER><FILEDESC><TITLESTMT><TITLE>South AfricanLiterary
Encyclop&aelig;dia</TITLE> </TITLESTMT><PUBLICATIONSTMT><P>Draft version of
page 1 only produced at CETH seminar 23 July
1996</P></PUBLICATIONSTMT><SOURCEDESC><P>This document based on ASCII
version of page 1 of hisory.doc "Concise historical survey: South African Literature
(incomplete) 1996. CSSALL, Private Bag X 54001, University of Durban-Westville, Durban,
4001</P></SOURCEDESC></FILEDESC></TEIHEADER>
<TEXT><FRONT><TITLEPAGE><TITLEPART>South African Literary
Encyclop&aelig;dia</TITLEPART><TITLEPART>CSSALL, 1996</TITLEPART>
<TITLEPART>Volume One</TITLEPART><TITLEPART>15th to 19th
Centuries</TITLEPART><DOCIMPRINT><PUBLISHER>CSSALL
Publishers</PUBLISHER><PUBPLACE>Durban</PUBPLACE><DOCDATE>1996</DOC
DATE></DOCIMPRINT></TITLEPAGE></FRONT>
<BODY>
<DIV0 TYPE="volume"><HEAD>Volume 1: 15th to 19th Centuries</HEAD>
<DIV1 TYPE="century"><HEAD>Part 1: 15th Century</HEAD>
<DIV2 TYPE="entry1bibl"><BIBL><DATE>1497</DATE></BIBL></DIV2> </DIV1>
<DIV1 TYPE="century"><HEAD>Part 2: 16th Century</HEAD>
<DIV2 TYPE="entry1bibl">
<BIBL><DATE>1552</DATE><AUTHOR>de Barros, Joano</AUTHOR> <TITLE>Da
Asia</TITLE></BIBL>
<DIV3 TYPE="entry2comment"><P>The first of four volumes appeared. It includes an
account of the Cape. De Barros (1496-1570) was the Portuguese Governor of St. Geore del
Mina on the West Coast from 1522 to 1525. He then became Treasurer of the Indian branch
of the Revenue, Councillor, and Historian.</P>
<DIV4 TYPE="entry3titlereview"><P></P>
```

```
<DIV5 TYPE="entry4keyword"><P>prose; portuguese; history</P><DIV6
TYPE="entry5fulltext"><P></P></DIV6></DIV5></DIV4></DIV3></DIV2></DIV1></DIV
0></BODY></TEXT></TEI.2>
```

Notwithstanding the advantages of SGML/TEI markup, it had by this stage in the project
become clear that the best strategy for the SALIT Web was a multi-layered approach in
which a bibliographic database could be used to "manage" SGML encoded full texts in the
Library section. Another option explored at this time, following my investigation of the
design of the *Mayibuye* CD-ROM, was the possibility of using a hypertext browser like
Microcosm or Folio Views for the SALIT Web as a whole. In addition to Pringle's *African*
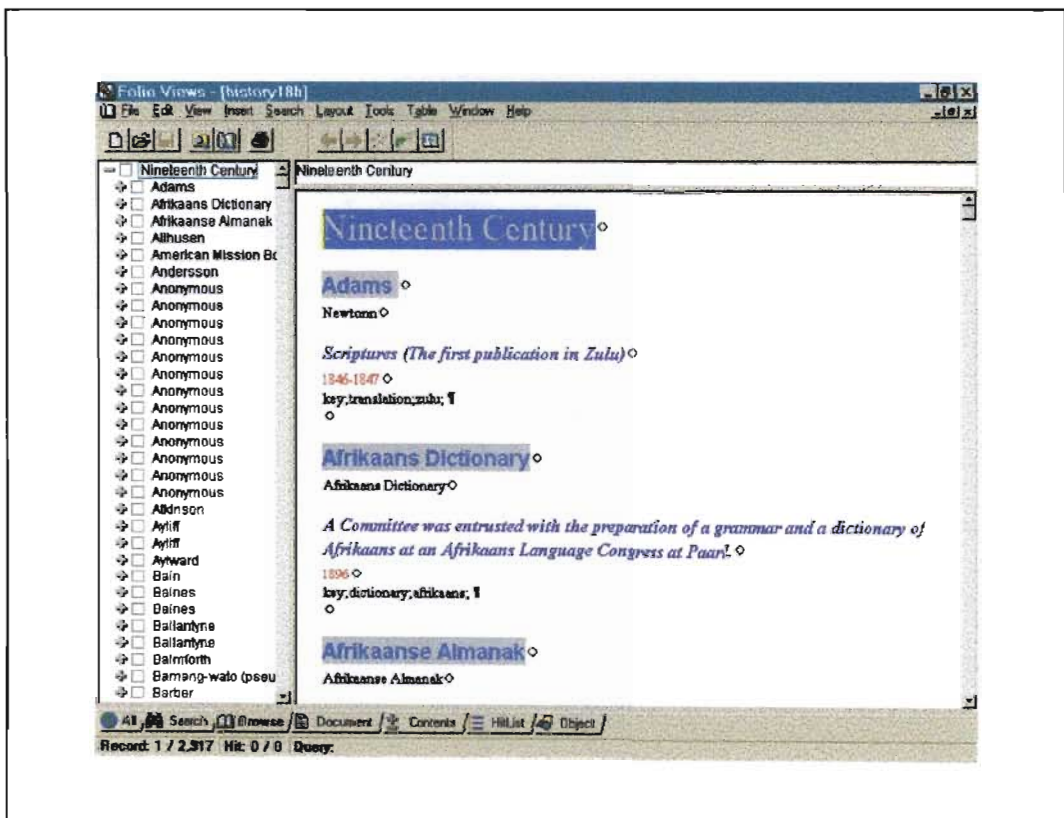
Figure 64: Experimental version of SALIT in Encyclopaedia form, using Folio Views. The
duplicate author entries were inherited from the old "working database".

Sketches (1834) and *Alternation* Vol 2:2 (see Chapter 4), I produced a pilot version of the database itself in Folio Views format, including hypertext links to the *African Sketches* full text (See Figure 64). While the decision to keep the MS Access database bibliographic structure within the SALIT Web eventually rendered this version redundant, the Folio Views texts are equally accessible via hyperlinks in the database. For the preparation of full-texts for inclusion in the SALIT Web, however, SGML still remains the best long-term strategy (See Chapter 6 for a discussion of the issue).

Chapter 11 Conclusion

In this thesis I have attempted to provide a comprehensive account of the development of the SALIT Web, framed by a discussion of current research into electronic text resources, and illustrated by the results of my own explorations of how they may be used for access, learning and research in South African literature. The interdisciplinary nature of this enquiry has uncovered relationships between literary studies and the formally separate domains of machine-readable text, computer-based concordancing and distributed learning systems, the direct outcome of which has been the design of the experimental SALIT Web CD-ROM; a design which integrates a bibliographic database, a virtual library of electronic texts and a multimedia encyclopaedia. In addition, the same resource can be used as a platform for learning, teaching and research, while providing the tools for textual analysis.

While the experimental CD-ROM is a tangible outcome of the present study, leading in due course to its formal publication as a multi-functional resource for South African literary investigation, it is hoped that the foregoing account of its development will have heuristic value by encouraging research into its future development both as an investigative tool, and as a networked repository for South African texts to be supplemented by and shared with other researchers in the field. A distinguishing feature of electronic text projects undertaken over the last decade has been their preoccupation with collaboration and the sharing of resources, and I have tried to show that the SALIT Web could initiate a similar pattern in South Africa where the communication infrastructure is now in place to support networked co-operative projects.

Ultimately, the value of my research lies not so much in bringing new knowledge to any one of the related disciplines on which it draws - literary studies, computing, learning or research methodology - but in revealing connections between them that have not previously

been perceived, and inter-relating them to support researchers, teachers and learners in the area of South African literary studies.

I was motivated to undertake this project while doing preliminary reading and research for an intended doctoral project on Anglo-Boer War journalists (war correspondents) who went on to write fiction with a South African setting. Finding information and then getting access to relevant material presented formidable obstacles. Although there were a few special collections, few of them yielded adequate information on marginal writers (with the exception of the National English Literary Museum (NELM), which provided some valuable assistance and material). Travelling the country to find additional material, I discovered to my alarm that one of the best collections was not registered with any publicly-accessible catalogue - I had come across it merely by chance, buried in the archives of a museum. This experience coincided with a growing personal interest in computing and using the Internet where I read an account of the work of the Text Encoding Initiative (TEI) (see Chapter 6), a wide-ranging research project aimed at standardising the format for electronic text in the humanities. The TEI and related literature-based hypertext projects promised the emergence of networked information resources that could absorb and then share the type of marginal texts I had been searching for. "Virtual libraries" already existed in Latin, Greek, French, English and American literature but there was no evidence of any attempt to collect South African literature in this way. By undertaking this project I hoped to start a process, based on international standards, that would provide a framework for a virtual library of South African literature, especially those works considered "marginal" or had gone out of print, or were difficult to access for a variety of reasons.

In Chapter 9, reference was made to the integration of modern electronic communication networks and conventional scholarly publications through the establishment of a virtual ongoing "seminar" with the potential to combine the best features of both. I

would argue that the pursuit of such a new medium for literary study, underpinned by an extensive corpus of primary texts in electronic form would be a highly productive direction for future development in South African literary studies.

Appendix A: Notes on installing the Beta 01 Version of the *SALIT Web* CD-ROM

The Beta 01 version of the SALIT Web CD-ROM that supplements this thesis is an experimental version not intended for distribution. While all the features described in earlier chapters are fully operational, Beta 01 is still essentially a "builders'" copy, requiring the pre-installation of several application programs on the user's computer for full access. Software requirements for pre-installation appear on the cover of the CD-ROM, but are repeated here for information.

Hardware requirements: a Pentium 1 (or higher) computer running Windows 95 or 98, and a CD-ROM drive. A sound card is necessary for the audio and video content.

The Beta 01 CD-ROM requires that the following programs be installed on the user's computer: Microsoft Office 97 (including MS Word, MS Access and PowerPoint); Folio Views 4.1; RealPlayer (available for free download from the Internet at http://www.real.com/products/tools/index.html) and TACT (also free - http://www.chass.utoronto.ca/cch/tact.html). The CD-ROM will not run without MS Access and MS Word and are therefore essential. However, not all the features will be available unless the other programs are installed.

Once these programs have been installed, place the CD-ROM in the drive, click "My Computer" and then the "SALIT" folder. Clicking "01 South African Literature CD-ROM.doc" will bring you to the Opening Page (See Figure 6 in Chapter 2).

Troubleshooting tips

● The Microsoft Office program's default installation contains automatic virus alerts whenever a new file is opened. The CD-ROM is virus-free so it is suggested that you temporarily disable this feature for ease of usage.

● TACT is a DOS-based program. Read the TACT installation guide for any problems that this might present on a Windows-based computer.

● Many database options that would normally be unavailable to the user have been left accessible for our own data input and re-design work. By following the guided structure in the SALIT Web you should not come across these options, but if you do, simply close the program and re-start.

Any further queries about the operation of the Beta 01 version should be directed to me at this email address: stewartg@pixie.udw.ac.za.

Appendix B: Conversion of the database from linear to relational form

In order to supplement the more conceptual account of the database conversion in Appendix A, the following step-by-step procedure has been included to guide future work that may involve incorporating single-table bibliographies into a relational database for use in a hypertext web. The applications used here were the Symantec Q&A Database for Windows Version 4 and MS Access 97.

Stage 1: Eliminating duplicate author names

1   Create new table from query: qbase_OLD (call it Title) using a make-table query

2   In query Design view, click the arrow next to the Query Type button on the toolbar, and then click Make Table. The Make Table dialog box appears.

3   In the Table Name box, enter the name of the table you want to create or replace (Title).

4   Click Current Database to put the new table in the currently open database (Ency1.mdb).

5   Click OK.

6   Drag from the field list to the query design grid the fields you want in the new table (all fields).

7   In the Criteria cell for the fields that you've dragged to the grid, type the criteria (these remain unchanged).

8   To preview the new table before you create it, click the View button on the toolbar. To return to query Design view and make changes or run the query, click the View button on the toolbar.

9   To create the new table, click Run on the toolbar.

10   To check, open the new table (Title) from the Database window.

Stage 2: Create a new Author table using "auto delete duplicate records from table" procedure

1   In this procedure, you create a copy of the structure of the table that contains duplicates (the new Title table), make primary keys of all the fields that contain duplicates, and then run an append query from the original table (Title) to the new table (Author). Because fields that are primary keys can't contain duplicate records, this procedure produces a table without duplicate records.

2   To create a new table

3   In the Database window, click the Tables tab.

4   Click the name of the table you want to delete duplicate records from (Title).

5   Click Copy on the toolbar.

6   Click Paste on the toolbar.

7   In the Paste Table As dialog box, type a name for the copied table (Author), click Structure Only, and then click OK.

8   Open the new table (Author) in Design view, and select the fields that contained duplicates in the table you copied (surname, first name).

9   Click Primary Key on the toolbar to create a primary key based on the selected fields.

10   Save and close the table.

11   To append only unique records to the new table, create a new query based on the original table containing duplicates (Title).

12   In query Design view, click the Query Type button on the toolbar, and then click Append Query.

13   In the Append dialog box, click the name of the new table (Author) from the Table Name list, and then click OK.

14   Include the required fields from the original table by dragging them to the query design grid (ID, Author1_ID, surname, first name, gender).

15   Click Run on the toolbar.

16   Click Yes when you receive the message that you're about to append rows.

17   Click Yes when you receive the message that Microsoft Access can't append all the records in the append query. This transfers only unique records to the new table (Author) and discards the duplicates.

18   To see the results, open the table (Author) from the Tables tab in the Database window.

Stage 3: Create relationships between the two new tables: Author and Title

1   Open the Relationships window.

2   Drag the two new tables - Author and Title - into the window.

3   Highlight "surname" and "first name" in the Author box.

4   Drag the highlighted items to the Title box.

5   In the create relationship dialogue box, make sure that "surname" and "first name" in both tables are listed alongside each other.

6   Choose option 3 as the link type (i.e. all Title entries to their equivalents in the Author table).

7   Delete old tables: Author1 and Title1.

8   Hide the qabase_OLD box.

9   Save the new relationships.

Stage 4: Add new fields to the Author table

1   From the Database window, open the Author table, and select Design View.

2   Add the following:

Field Name	Data Type
Honorific (Rev, Dr, Ms etc.)	Text
Pseudonym	Text
Biography	Memo
Birth date	Text
Death date	Text
Picture	OLE Object
Sound	OLE Object
Reviews	Hyperlink

3   Save the table.

Stage 5: Add new fields to the Title table

1   From the Database window, open the Title table, and select Design View.

2   Add the following:

Field Name	Data Type
Commentary	*Change to* Memo
Full text	Hyperlink

3   Save the table.

Stage 6: Create new table: Year

1   Choose Get External Data from the File menu.

2   Import numbers2.txt.

3   In the Import Text Wizard dialogue box, choose "advanced" and change Width from 3 to 4.

4   Continue through the Wizard, choosing Create New Table, and changing Data Type from "long integer" to "text".

5   Save the new table as "Year".

Stage 7: Create new table: Events

1   In the Database Window, choose new table.Use Wizard to create new table, selecting "Events" sample from MS Access.

2   Retain only the following fields:

3   Event ID

4   Event Name

5   Event Type ID

6   EventDate (change from StartDate)

7   Event Description

8   Save table as "Events".

Stage 8: Create relationships between new tables
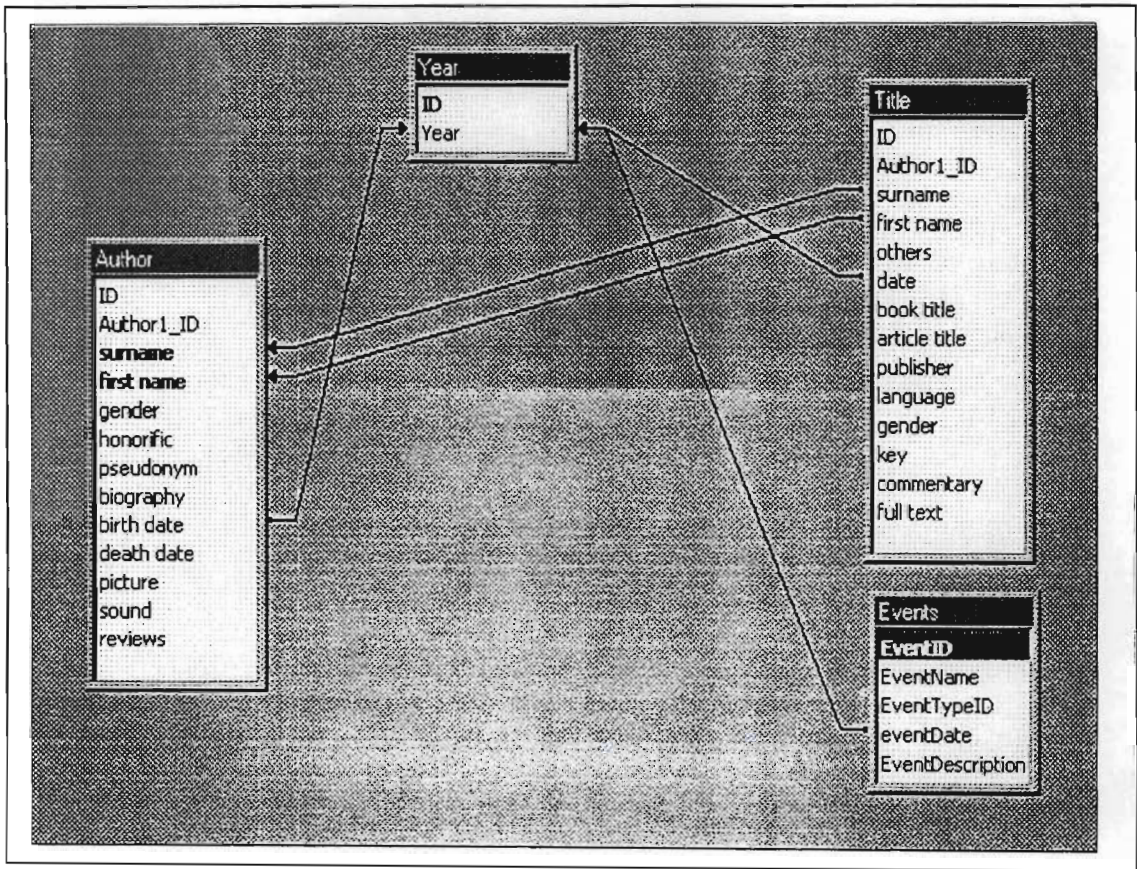
Figure 65: Relationships between joined tables

Stage 9: Delete old Forms and create new ones to correspond to the new Tables

1   Create new forms for

        Author:

        Author (Tabular View - Colourful 1)

        Author Subform (Datasheet View - Standard)

        Title:

        Title (Datasheet View - Standard)

Events:

Events (Datasheet View - Standard)

Year:

Year (Tabular View - Colourful 2)

2   Edit Author form to include Title as a sub-form.

3   Edit Year form to include 3 sub-forms: Author Subform, Title and Events.

Stage 10: Create coherent directory structure for linked and hypertext objects

1   Planning the final CD-ROM requires that data and programmes be contained within a coherent directory structure in the C drive. The structure will ensure that link integrity is maintained from one computer to the other, and on the CD-ROM.
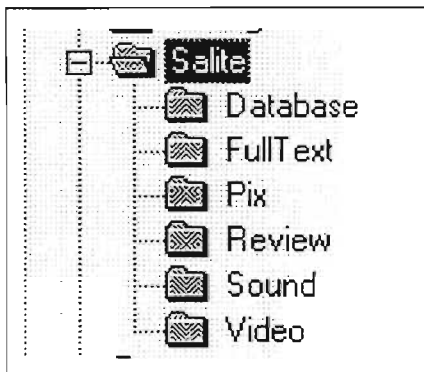
2   Structure map:

Figure 65: Directory structure

The structure can be expanded to include sub-folders, e.g. under Full Text, folders for Folio Infobases, Word Documents, etc.

Reviews could be similarly expanded, if the documents are saved in several different formats (although it would be preferable to stick to Word).

Sounds could be expanded to include *.wav files and *.ra files.

3   Where possible, compressed files should be used, e.g. JPEG for photographs and pictures, RealAudio for sound and video files. RealAudio still needs to be registered within MS Access.

Appendix C: Producing Electronic Texts: A Practical Course Outline

The SALIT resource has been developed in the context of two academic courses at South African Higher Education institutions: the South African Literature and Languages post-graduate courses at the Centre for the Study of Southern African Literature and Languages (CSSALL) at the University of Durban-Westville and the Department of Library and Information Studies at the M.L. Sultan Technikon. Collaborative projects involving students from both courses have been planned during the development of SALIT, and the first of these, an electronic Web with a focus on the work of John Langalibalele Dube has already commenced. To remain sustainable, the SALIT project requires a long-term research and development plan that will draw on the unique skills of students from both the disciplines (Literary Criticism and Library Studies). Such a plan would ensure the continuing development of the resource while simultaneously broadening the user base of the SALIT Web both by learners and researchers. Although the Dube Web project is at the time of writing still in a very early stage, the project design already suggests capacity-building in the creation of electronic texts as a crucial area of skills development for both Literature and Library Studies students. The project also comes at a time when South African Higher Education, and in particular the Humanities, needs to use information technology to match current practice in world-wide academic research.

> The development of well-resourced libraries and state of the art information technology is pivotal for the effective functioning of the South African higher education system in the information age and as part of the global academic community. (*Green Paper on Higher Education Transformation* Appendix 1 December 1996)

The B. Tech: Library and Information Studies includes the study of Information Retrieval and Electronic Texts and aims at building the capacity required to meet the objectives described in the Green Paper, while the University of Durban-Westville project aims at the

publication/networking of a major full-text database. Thus the potential exists for staff and students at both institutions to participate in a mutually beneficial programme to build research skills, and provide training at a high level in the capture, digitising and tagging (encoding) of electronic text. The technikon is increasingly being asked to provide trainees to convert existing hard-copy records into digitised form for inter- and intranet purposes.

Under the supervision of staff from both institutions, the Co-operative project consists of several smaller projects involving pairs of students from the two institutions under the rubric: *South African Literary Texts in Electronic Form.* The division of roles within the collaboration is roughly as follows: the UDW contribution includes the research, discovery, location and re-evaluation of Southern African texts (in various languages both written and oral); while the M.L. Sultan Technikon contribution includes classification (design of a scheme that matches the knowledge structure; creation of an authority list), design of the hypertext link structure; digitisation, encoding and preparation of the texts for electronic publication.

The result of the collaboration would be valuable, accurately encoded electronic texts of known provenance that would contribute towards the building of SALIT as a national literary resource as well as gain qualifications and build the research capacity of the participating technikon and university students.

The following table summarises the roles of each of the partners in the Dube Web project plan:

CSSALL (UDW)	MLST
Supplies dissertation	Knowledge structure design scheme; Hypertext design
Monographs	Preparation of word-processed documents for Folio Views
Translations	Scanning & OCR of non word-processed materials
Other primary texts	Proof-reading of all textual materials
Editorial Consultant	Scanning of archive material and photographs
Location of:	Recording and digital transfer of sound material to RealAudio files

CSSALL (UDW)	MLST
Photographs	Locating and contact with copyright holders
Sound/ audio material	Applying for copyright permissions
Archive material	Applying for rights/ license agreements for use of the following software:
Newspaper articles	MS Word
	Folio Views
	RealAudio

Many of the skills developed within the present project, including the epistemology of a Literature Web and the practical skills of electronic text preparation, hypermedia and computer-based textual analysis would have to be cascaded down to students working towards the creation and use of the resource. The following course outline has been developed as the basis for a pilot project to be run jointly by the two institutions. Much of the physical infrastructure (computer networks, scanning equipment, etc.) is now in place, and researchers and staff members at the institutions have the practical experience to facilitate the learning. It is likely that visiting academics over the next two to three years will assist in developing the course.

The course outline below is expressed as an outcomes-based unit standard consistent with its inclusion as a modular component in either of the courses involved in the present collaboration, or as a stand-alone offering.

UNIT STANDARD		
		Page of
Preparation of Electronic Texts		
Unit number: ETEXT01	**Field**:	Literary Criticism;
Level: 6		Library & Information
Credit: 20		Studies
	Issue date:	
	Expiry date:	

Capability: A candidate at this level is capable of designing and creating an electronic text resource.

Entry Assumptions:
(a) 3-year degree/ diploma (Level 5) qualification in a Humanities discipline.
(b) Word processing skills at Level 1.

Performance Outcomes:
1. Identify document types, and apply appropriate markup schemes.
2. Scan, encode (mark-up) and insert metatextual information into an electronic text.
3. Digitise graphics, audio and video files for integration with an electronic text resource.
4. Select and implement a suitable user interface for an electronic text resource.

UNIT STANDARD

Assessment Criteria:

1.1 Document types (literary, business and user manuals) are identified, and their elements described.

1.2 The history of the development of electronic texts is described.

1.3 System-independent text formats (ASCII, SGML, HTML) are distinguished from proprietary formats (PDF, Word-processors) and their applicability to the uses of electronic texts evaluated.

1.4 Digital imaging formats are described in terms of resolution, compression and suitability for integration into electronic texts.

1.5 Digital audio and video formats are described in terms of quality, compression and suitability for integration into electronic texts.

2.1 Documents are scanned and processed using Optical Character Recognition (OCR).

2.2 Documents are assembled, proofed and edited.

2.3 Documents are encoded in Word Processor, HTML and SGML formats.

2.4 Images are digitised, saved in various formats and integrated into electronic texts.

2.5 Audio and video files are digitised, saved in various formats and integrated into electronic texts.

3.1 User interfaces are described (hypertext viewers, including Internet browsers and proprietary text database environments).

3.2 User interfaces are applied to electronic text resources (Word Processor, HTML and SGML formats).

UNIT STANDARD
Page of

Underpinning Knowledge:

Performers must be able to:

i.   Identify the stylistic features of major literary genres as well as business documents and user manuals.

ii. Operate a computer and commonly used Word Processing application programs.

iii. Apply knowledge classification schemes and the basic techniques of information retrieval.

Range Statements:

1. The electronic texts prepared should meet the requirements in terms of textual accuracy and appropriateness of the markup for their intended use.

2. Theoretical knowledge should be assessed in the context of the activity and the standard of the output (not simply as a written or oral test).

3. The electronic text interface should meet the required standard in terms of user accessibility, relevance to the knowledge structure of the document/s, and care in visual design features.

Moderation:

This unit to be moderated by staff of the CSSALL, the Department of Library and Information Studies and a professional non-academic user, e.g. from the Public Library system or an Electronic Text Centre.

Expressed as a 30 hour programme (60 notional hours, including research and individual practical project development), the course could be offered as follows:

Course Content		Hours
3	Background	2
3.1	History of e-text (OTA etc.)	
3.2	Differing platforms & uses	
4	Identification of texts	2
4.1	Literary texts	
4.2	Business documents	
4.3	Manuals	
5	Origination of e-texts	2
5.1	Ready-made commercial titles	
5.2	Internet	
5.3	Typing	
5.4	Scanning & OCR	
6	Practical e-text preparation (see above)	6
7	Structured documents and mark-up (encoding)	6
7.1	ASCII	
7.2	Word processors	
7.3	SGML (HTML, TEI)	
8	Digital imaging	6
9	Viewers and hypertext	6
9.1	Internet browsers	
9.2	FolioViews	
Total Hours		30

Appendix D: A home for our marginalised past: Creating a virtual library of South African literature

This is a copy of a paper presented at the South African English Academy Conference "English at the turn of the Millennium", 16 September 1998.

Abstract

Although the entire spectrum of South African literature is now potentially available to all South Africans, there are new formidable obstacles to realising this ideal. In place of censorship we have impoverished libraries, rationalised school and university budgets and the rising cost of bound and printed books. While South Africans are building an open society the printed word is threatened globally. "The displacement of the page by the screen is not yet total ..." writes Sven Birkerts, "but the large-scale tendency in that direction has to be obvious to anyone who looks." (*Gutenberg Elegies.* 1997: 3). This paper describes a collaborative project between the University of Durban-Westville and the M.L. Sultan Technikon to produce electronic texts on CD-ROM and the Internet as cheap and accessible alternatives to print publication, but also for analysis, discovery and learning. Individual projects under way are: Thomas Pringle's 1834 edition of *African Sketches*, the life and works of John Langalibalele Dube and Internet access to a South African Literature database. Collaboration amongst South African educational institutions could ensure that rare, marginalised, or out-of-print texts are encoded in ways which will make them easily accessible to scholars, thus building our computer-based literary and linguistic corpus.

Electronic texts

Although I have quoted Sven Birkerts' comments on the increasingly pervasive presence of computer-based text in our lives in support of my argument that electronic text is the saviour of South African literary heritage, I must admit at the outset that I have used his words a little tendentiously. Birkerts' attitude towards electronic text as expressed in *The Gutenberg*

Elegies is unambiguously negative. Like many of us who love reading and for whom literature has a special appeal, he expresses regret at the demise of the printed book, even the sensual appeal of the book as a comfortable, practical and effective medium, easy to hold and ideally suited for the act of reading. More significantly, he has profound doubts about the suitability of screen-based text for conveying coherent thought. I recommend his chapter "Hypertext: Of Mouse and Man" to anyone who has valiantly attempted to read hypertext fiction. But Birkerts' *Elegies* are grounded in a privileged Western sense of a *fin de siècle* collapse of familiar academic environments that, unlike our own in South Africa, can still afford well-stocked libraries and can choose to ignore, for the time being, the challenge that the new media represents for the preservation of literary heritage and in particular, the voices of the disenfranchised.

I make no apologies for being considerably more up-beat than Birkerts in my attitude towards electronic texts, and while I accept that dangers lurk in the indiscriminate acclamation of computerised books, I also believe that literature must stake its claim in the apparently unsympathetic realm of Information Technology. For most South Africans, this country qualifies as one of the world's "bookless places" as this state has been described in at least one debate on electronic texts (Hart et al.1998). Apart from the economic impossibility of building or maintaining extensive book-based libraries in our new schools or universities, we have a vast untapped resource of previously marginalised literature - novels, poetry, drama, diaries, historical accounts and travel writing - that were denied to us by political or commercial proscriptions. Ironically, at a time when our newly open society is ready to rediscover previously unavailable texts, the expense of book production prevents it. Our library shelves will continue to reflect the narrow collection policies of the past because we simply cannot afford to redress the situation; and even if we could, the titles would often not be there to buy.

Although the specific missing texts might be unique to South Africa, the phenomenon of inaccessibility to marginalised writing is not. Women's writing for example, has been similarly subjected to political, commercial and societal suppression worldwide. Two major electronic texts projects, the Brown University Women Writers Project (1996), and the Victorian Women Writers Project have sought to address the historical devaluing of women's writing by re-publishing or making available for the first time, works that are essential for a full appreciation of women's place in literature. The same impulse, but in the context of South African state oppression, motivated the development of the Mayibuye Centre's *Apartheid and the History of the Struggle for Freedom in South Africa* CD-ROM.

Electronic text projects and marginal writing

Brown University Women Writers Project

The original objective of the project was to produce a single scholarly anthology that asked new questions about genre, publication, canonical traditions and literary culture. The strong literary computing legacy at Brown University turned the founders of the project to computer technology as a natural tool for providing efficient and innovative access to this material. As the project developed, its aims were expressed as: (i) to produce scholarly anthologies and individual volumes of pre-Victorian Women Writers; (ii) to contribute to research on literary databases and text management systems; and (iii) to serve as a central research resource with a full-text electronic database of pre-Victorian women writers in the English language. Landow's previously-established *Victorian Web* (1989) at Brown University had laid a foundation for the literary orientation of the computer technology there.

The Women Writers Project is an important model for the development of literary "textbases" in South Africa for two reasons: firstly, by building an electronic collection of women's writing in English between 1330 and 1830, material which has been largely inaccessible to researchers and students is becoming available for study; and secondly, by making previously marginalised texts available, the project is able to challenge

misrepresentations of the role of women in Western literary and cultural history. The parallels with the suppressed and marginalised literature of our South African past are clear, and the Women Writers Project offers a practical way in which to collect and make accessible important texts that have long been out of print, or perhaps never formally published. The University of the Western Cape's Mayibuye CD-ROM project has already had success in this respect. Where the Women Writers Project textbase disseminates its texts through on-demand publishing, traditional print publishing of selected works, provision of selected texts to database publishers, and Internet access, the Mayibuye project has distributed print copies of the works contained on the CD-ROM as part of its "library" package.

The Women Writers Project, in contrast to Landow's work in the *Victorian Web* (formerly *Context32*) (1989: 173-198), has from the start committed itself to the encoding of its texts in SGML (Standard Generalised Markup Language). Mah and Flanders (1996: 1) explain that SGML allows the texts to be encoded to help the user to navigate within the document, analyse it, compare it to others and find words and concepts within particular contexts (e.g. speeches, quotations, or notes). The TEI (Text Encoding Initiative) (1994) SGML encoding system for humanities texts also contains tags that allow parallel versions of the text: one with the original typographic and other errors (that may be important for research purposes) and another more readable "corrected" text. The careful coding of texts has led to the Women Writers Project becoming an important scholarly resource in the field of pre-Victorian women's literature.

Victorian Women Writers Project

The Indiana University based Victorian Women Writers Project (1996) also aims to increase accessibility to works previously difficult or impossible to obtain. Concentrating on the writing of British women from the Victorian period (1830-1910), the Victorian Women Writers Project began offering open access to the re-published works through the Internet in

1995. The project has ensured that writings by the selected group of seventeen late-19th century British women authors are now receiving more exposure than they received from their contemporaries. This project, like that at Brown University, constitutes a "virtual library" offering a comprehensive specialised collection that could not be matched by a conventional print-based library. The model developed for both these projects - combining a library-orientated special collection goal with textual and literary research - has helped shape the project under way at the Centre for the Study of Southern African Literature and Languages (CSSALL) at the University of Durban-Westville. Attention to the encoding of the texts (e.g. in accordance with the TEI guidelines) ensures that they can measure up to scholarly publishing standards and can be used by researchers in the confidence that they are in nearly all respects (except that of print publication) new editions of known provenance. A cautionary note is necessary in considering electronic texts available on the Internet or CD-ROM because there are titles available that lack careful attention to accuracy and omit important information on the preparation and editing of the texts. Project Gutenberg, for instance, has the admirable goal of making 10 000 books available in their "virtual library" by 2001, but in order to achieve this figure, its promoters explicitly reject complete accuracy and a digital format that would require an account of the editorial history of the text (1992). Any South African endeavours in the creation of virtual libraries would do well to follow the route taken at Brown and Indiana Universities.

Mayibuye CD-ROM

Although its texts contain no descriptive markup (e.g. SGML) , the books published on the Mayibuye Centre's CD-ROM *Apartheid and the History of the Struggle for Freedom in South Africa* represent the most successful achievement in a South African context of establishing an electronic text resource. The aim of the Centre is to "help recover areas of South African history which were neglected in the past and to create space for cultural creativity and expression in a way that promotes a process of change and reconstruction in a democratic South Africa." The CSSALL's SALIT Web project would endorse this aim, but

aims to be more inclusive, extending its "library" to texts beyond those specifically related to South Africa's liberation struggle. Included on the Mayibuye CD-ROM are the full texts of 50 books published between 1978 and 1994, including biographies, poetry, and novels (e.g. Mary Benson's *South Africa: The Struggle for a Birthright* and Alex La Guma's *And a Threefold Cord*). Like Mayibuye, the CSSALL's SALIT Web project includes multimedia material in addition to the electronic texts. Electronic publishing lends itself to multimedia presentation and can integrate photographic, audio and film/video along with any texts it may contain. We are already familiar with the integrated text, graphics and sound files to be found in CD-ROM encyclopaedias. Turned to literary purposes, these features can provide an immensely rich resource, as the Mayibuye publication or Landow's *Victorian Web* are evidence.

Other South African projects worth mentioning are the UCT Poetry Web and the UNISA Electronic Text Centre, that while different in intention, share the potential for providing wider access to South African texts. Links to these and other electronic text resources and projects are available as a supplement this paper on the Internet. The page may be accessed via the University of Durban-Westville website (follow the links to the CSSALL) or directly at this address: http://www.udw.ac.za:80/~stewartg/virtual.html. At this site you may, for example, browse through the works available at Brown University, and perform online searches through the full texts. I would particularly recommend using the University of Virginia's E-Text Center (search the modern English collection) (http://etext.lib.virginia.edu/) and R.J.C. Watt's *The Web Concordances* (English Romantic Poetry) at the University of Dundee (http://www.dundee.ac.uk/English/wics/wics.htm).

UDW CSSALL SALIT Web project

Since the beginning of 1996 I have been involved in a project at the CSSALL aimed at the publication of an electronic literary resource called SALIT on CD-ROM. Based on Johan van Wyk's *Concise Historical Survey: South African Literature* (1996b) and originally

conceived as a support structure for the *South African Literary Encyclopaedia* project - also located at the CSSALL - it has grown into a separate publication in its own right. It shares some of the aims of the international virtual libraries mentioned earlier, but is also intended as a multimedia publication similar to the Mayibuye project.

The SALIT Web operates as a literary resource in several ways. It can be used in all of the following ways:

As a bibliographic database

As a virtual library

As a multimedia encyclopaedia

As an archive for the preservation of marginalised texts

As a textual analysis tool

As a learning and teaching programme

During the course of the project the three members of the research team have surveyed the literature and computer programs relating to these different functions, experimented with the design, tested various types of software and evaluated their appropriateness to the needs of a range of learners and researchers in South African literature and literary history. In addition, the team has scanned (produced digital copies) of a number of significant texts and begun the process of encoding these in ways that would make them more readily accessible to literary researchers. One of these texts, Thomas Pringle's *African Sketches* (1834) is discussed more fully later. The development of the database itself has been fundamental to all these activities, and at the present count contains references (with varying degrees of additional information) to 8438 authors, and 31,889 titles and events extensively cross-referenced and indexed in several ways including gender, language, genre and theme. The multimedia content includes a gallery of 507 photographs and a small but growing number of audio and video clips.

While all the uses of SALIT Web listed above have significant implications for learning, teaching and research in South African literary studies, two distinct categories of application have emerged: general access to information (with features similar to bibliographic databases, virtual libraries, multimedia encyclopaedias and archives for the preservation of marginalised texts), and research and teaching methodologies (textual analysis tools and learning and teaching programmes). A further intriguing category beyond the scope of my present study but one that invites further investigation is the use of electronic text (in particular hypertext) for creative or imaginative composition, Sven Birkerts' reservations notwithstanding.

As it is the "virtual library" function of the SALIT Web that concerns us here, I will go on to provide an account of two of the electronic texts that are contained in its web of resources.

Pringle's *African Sketches* (1834)

Its publication in the SALIT Web restores the original format of the work: combining Part 1: *Poems Illustrative of South Africa* and Part 2: *Narrative of a Residence in South Africa* into a single electronic text. A full word index allows quick searching for words and phrases and allows cross-referencing of poems, notes and the text of the *Narrative*.

Thomas Pringle's 1834 text *African Sketches* was selected as a suitable pilot project in which a significant South African text could be located at the centre of a hypertext web. In addition to its being one of 31,889 "title" references in the bibliographic database of the SALIT Web, it was intended as an example of how each SALIT reference had the potential, over time, to be linked to the full text to which it referred. The design of the electronic publication followed as closely as possible the original print-based book - cover pages and front matter, Parts 1 and 2 of the text, the contents, major section divisions, footnotes and plates. The database provides a hypertext link to the title, and from there on, each of the elements listed above are hyperlinked within the micro environment of the electronic web structure.

The Pringle text offered considerable scope for the development of a wider information web around the original text, including other works by Pringle, contemporary documents of historical and social significance, audio and photographic information, maps, commentaries, reviews and criticism. Another promising avenue for the assemblage of an information web was Pringle's involvement with the establishment of South Africa's first newspaper, *The South African Commercial Advertiser* and the subsequent bitter conflict between the publishers and the governor of the Cape, Lord Charles Somerset.

The process of producing the e-text - starting with the origination of the text (scanning) through later stages of proofreading, editing and setting within a hypertext environment - proved more complex and time-consuming than anticipated. The complexity resided not only in the technical challenges posed by the digitising of the various documents and visual material, but also by the conceptual challenge of valid electronic representation of textual and meta-textual elements (like authorial footnotes and biographical material), developing search criteria, classifying and identifying hypertext links (for example, discriminating between simple explanatory pop-ups containing later critical commentary or the author's own footnotes and endnotes). The relationships amongst these various meta-textual elements now present themselves as a clearly-defined system to the reader/user and although some of the more familiar relationships have an established place with reference to the original text - like the footnotes and endnotes - others do not. For instance, *African Sketches* has a chapter consisting entirely of biographical notes on some of the personalities referred to in the body of the text. The scope of these biographies is constrained by the 1834 publication date, and fuller details (including the date of death, for example) required clear demarcation as later editorial additions. The most appropriate way in which to incorporate these elements had to be considered within the conventions of scholarly publication while taking into account factors based on hypertext theory such as ease of navigation and coherence in an information web,.

The present electronic version of the *African Sketches* text allows the reader the same access to the material as her 1834 peer: i.e. all the poems in the *Poems Illustrative of South Africa* section and the *Narrative of a Residence in South Africa*, together with all authorial notes, footnotes and endnote material. Beyond the more obvious differences in the medium that the present-day reader of the electronic version would use (the computer screen rather than the printed page) there are more profound differences. The electronic version of *African Sketches* can be read like any conventional book, but may also be used as an interactive search and discovery medium. Should the reader wish to enquire about the nature and quantity of Pringle's references to women, particularly black women, she could perform several searches - for example for the occurrence of the word "woman" within five words of the word "black" throughout the 115 poems and 350 pages of the two texts - and this would take less than 30 seconds. The results of such analysis form the basis of a separate paper, still work in progress.

Because additional material has been added to the broader environment of the database, detailed maps and photographs of the Eastern Cape and the Baviaans River Valley are also available to the interested reader. The journalistic possibilities around the *South African Commercial Advertiser* theme have not yet been developed, but facsimile copies of some pages of the paper have been included as well as contemporary illustrations and documents related to Pringle's own account of his conflict with Somerset in Chapter 10 of *Narrative of a Residence in South Africa*.

The Folio Views software used to view and navigate through the *African Sketches* text is designed to encourage interaction between reader and text, not only through hypertext linking and speedy search and retrieval, but through being able to add one's own annotations and material to the texts (either temporarily or as a permanent addition to the electronic text). This option opens up considerable scope for using the *African Sketches* text for further research and for learning. A scholar exploring Pringle's references to, say, "work" and

"economy" would be able to add permanent links to the text, as well as incorporate her own notes, commentary and related articles to the original web of information. In a very real sense, the reader/learner becomes a co-editor of the text and the centre of another web of references customised for the use of another set of researchers or learners.

The Dube Web

In a recent collaborative project between the CSSALL and the Department of Library and Information Studies at the M.L. Sultan Technikon, translations of the writing of John Langalibalele Dube, early 20th Century politician and founder of *Ilanga Lase Natal,* form the basis of an electronic "web" that will eventually contain a biographical monograph together with Dube's original work both in the original isiZulu and in English translation. The M.L. Sultan department specialises in the creation and retrieval of electronic texts, and are therefore ideal partners in literary research based at the CSSALL both through the contribution of technical information technology expertise, and in exploring the knowledge structure of the material for complementary hypertext design.

Conclusion

Researchers in European or American literature already have access to extensive textbases that they may utilise and may or may not overtly acknowledge in their work (is it necessary to state explicitly that alongside your dog-eared Jane Austen novels you happen to have the electronic versions, with every word indexed in several ways and extensively cross referenced?). Or that in your study of Dryden, you are able at the touch of a button to consult an electronic reference that provides you with every other occurrence of that word or phrase in English poetry from 400 to 1900? South African writing in English is barely represented in the world's virtual libraries, and its literature in other South African languages non-existent. Others have proved it possible to overcome the obstacles of incompatible computer files, copyright permissions and clumsy user interfaces to make literary texts available via the computer. I hope that the examples of the South African projects discussed

here will encourage more collaborative work aimed at building an indigenous electronic literary corpus that may lead to the establishment of a networked cultural heritage accessible to all.

M.L. Sultan Technikon, August 1998

References

(See main References section)

Appendix E: Evaluation of application programmes for displaying and searching electronic text

Decisions on "data" and "software" were central to this study, in order to fulfill the main aims as expressed in the title: to investigate "the implications of e-text resource development for Southern African literary studies in terms of analysis and methodology" and to implement these findings in the form of "an e-text database, and experimental CD-ROM". A combination of the review of the literature, direct experimentation with different software packages, interviews with developers and publishers of electronic text projects, and the standards adopted by centralised data services such as the ADHS (see Chapter 7) have led to the conclusion that the ideal "data" decision - i.e. the encoding of the electronic text - is the adoption of the TEI tagset of SGML.

Arguably the most crucial decision made during the course of the present study was the choice of a "front end" or *representation* system for the two hypertext projects identified as the outcomes of the study: the *Encyclopaedia* and Pringle's *African Sketches*. In very practical terms, this choice depended on an evaluation of the literature and software to supply a solution that was theoretically sound, appropriate to South Africa (where any such resource would be primarily used) and affordable. As the hypertext projects were to be available on CD-ROM and on the Internet, the "front end" would have to be user-friendly in both these media. Central to the decision was the extent to which the hypertext projects would be useful platforms for research (and for learning).

Making a decision on the "front end" application was also inextricably linked to the preparation (or encoding) of the material as well as the ease of access, both for conventional reading or browsing, and for analysis. Although the ideal "software" decision would have been the use of a powerful "Viewer" like Dynatext, in the context of South African

education, however, considerations such as cost and availability of skilled text encoders impact strongly on the feasibility of such a project.

Although the process of evaluating hypertext systems was anticipated in the research design, the recursive nature of the activity required frequent revisions of strategy, with consequent impact on data capture and the specific "shape" of the hypertext projects themselves. The value of adopting such a flexible plan of action was the truly organic development of the results, shaped as they were by the three major variables: the encoding of the text, the representation of the hypertext, and its suitability to South African literary studies (see Chapter 10 for a fuller account).

Keeping the options open on a final choice of "front end" was an inevitable concomitant of the research strategy but led to a number of pilot schemes that were also to prove blind alleys. Very early in the project, the value of keeping the data input close to an ASCII common denominator arose in the conversion of the *Concise Historical Survey* (Van Wyk 1996b) document from MS Word format into a Q&A database. At several points in the project, conversions to ASCII enabled the transfer of large and valuable parts of the project from one platform to the other. Even towards the end of the project, when final decisions on software platforms had eventually been made, ASCII proved helpful in setting up the pilot version of the on-line Internet version of the *Encyclopaedia* database. Beyond most of the considerations in the discussion of hypertext management that follows, the transfer of the material remains the single most vital element - underscoring a critical factor that instigated the development of SGML and the TEI: the portability of the text. It is likely that the same characteristic will ensure that HTML becomes the basis of future word-processing systems, ensuring seamless transfer of material from the "typist" to the Web.

To trace the process of reaching a decision on the most appropriate hypermedia "front end" it is helpful to re-visit our definition of the term. Hypermedia consists of nodes, links,

anchors (+ destinations) and networks; also it is *interactive*. Let us examine each of these terms more closely.

Nodes are discrete chunks of information (e.g. an Internet "page") and constitute the basic units of information storage. Because the information is modularised into essentially "independent" nodes within a hypertext environment, it is the user who determines the order in which information is accessed, and not the system. The "granularity" or size of nodes can vary enormously, but the smaller they are, the quicker they can be processed. Links in hypermedia demarcate associative relationships between the nodes that they join: a link joins a start point to an end point. These start and end points are called anchors, and consist of references to all or part of a node.

The organisation of information into separate nodes and connected by links is generally known as a "network". Such networks can be relatively circumscribed like the Pringle CD-ROM and can be described as a "hyperdocument" or a "hypertext project". The Internet is an example of a much wider network. Within a network, interrelated and interconnected ideas are linked together on the basis of a blend of semantic relationships and views imposed by the user so producing the information model of the knowledge available. Importantly, nodes and links can be grouped together into abstractions or composites that act as super-nodes within the information framework.

Hall (1996: 8) identifies three distinct methods of representing and storing hypermedia: (a) embedding the link in the node data, (b) the hyperbase approach and (c) the open link service approach. Landow (1994: 3) concurs with Hall and other commentators that the first method (typified by Apple's HyperCard 1.0) used by the World Wide Web and Windows Help is not "true" hypertext with limitations that prevent the user from exploiting the full capacity of the medium. One of the advantages of the system, the ease with which links may be edited and nodes moved around (maintaining outside links) also encourages "spaghetti" structures and

dangling links (the familiar "link has moved" message on the Internet). Even though the Web provides interconnected information, it is not a coherent or organised structure for publication; users are forced to "surf the net" repetitiously in search of information in what Hall calls "a model of information tourism" (Hall 1996: 3). The embedded link systems also tend to have relatively inflexible nodes that restrict the size of pages and force exaggerated "chunking of material with poor navigability within each chunk. The software is more suitable for electronic address books than literary texts.

The remaining two hypertext architectures, (b) the hyperbase approach and (c) the open link service approach are the most suitable for the literary projects envisaged as part of this present dissertation. By using the MS Access database to manage the contents of the SALIT Web, we have adopted an adaptation of the open link service, keeping our raw data (the full texts) in portable formats - MS Word/ASCII.

The basic computer equipment and software necessary to begin using electronic texts for teaching and research in literary studies are a computer, a flat-bed scanner, a relational database, and software suitable for SGML markup and viewing/browsing. In addition, a text analysis program such as TACT would allow the analysis of marked up text.

The following table summarises some of the factors influencing a decision on strategies for the development of electronic text collections.

	Folio	Dynatext	Microcosm	Netscape
Hypertext	√	√	√	√
Graphics	√	√	√	√
Audio/video	√	√	√	√
Auto categories	✕	√	✕	✕
Contents page generation	√	√	√	√
Automatic indexing	√	√	√	✕
Accepts multiple file formats	√	✕	√	✕
Boolean search	√	√	√	✕
Proximity search	√	√	√	✕
Internet capabilities	√	√	✕	√
Access management	√	√	√	✕
Non-proprietary ("open") encoding	✕	√	√	√
Level of encoding skill required	low	high	low	medium
Cost: builder/editor	R4,000	R12,000	R6,500	R200
Cost: Viewer	R250	R350	R100	R200
South African Users	HSRC; Jutastat	None	None	Majority of internet users

Figure 65: Development strategy /software matrix

During the course of the project, the following additional editing/viewing software was used and evaluated:

Softquad Hotmetal

(HTML editor - used to design the internet home pages for the Centre for the Study of Southern African Literature and Languages - see Chapter 6). Although the Hotmetal

software is reliable and flexible, for non-professional users it has been superseded by HTML publishing options in popular word processing packages (WordPerfect and MS Word).

WordPerfect 7 & 8

(HTML publishing option). This was used for the design of all additional HTML files on the CSSALL home pages. MS Word 97 and MS Frontpage were later found to be much more flexible, and were used for all subsequent HTML documents.

SGML Tagwizard

(HTML/SGML editor - installed and run from within MS Word 6). This shareware program proved to be useful and versatile for editing and especially converting HTML files. It was used extensively at all stages of the project. The "strip tags" feature, allowing for the conversion of HTML documents to uncluttered text files is not available in the more expensive commercial packages reviewed. Superseded by MS Word 97 and MS Frontpage.

Softquad Author/Editor

(SGML editor). The researcher was able to develop a preliminary draft SGML structure for the *Encyclopaedia* while attending the CETH Summer Seminar at Princeton University (see an account of this in Chapter 10). WordPerfect 8 now includes an SGML authoring feature.

Softquad Panorama Pro

(SGML Viewer). Sample versions of Panorama were distributed free to internet users by Softquad to promote the use of SGML on the World Wide Web. The "Pro" version was acquired by the CSSALL together with the Author/Editor package. Although it exploits the

considerable advantages of SGML markup, the search facilities require the input of specific tags that are not immediately apparent to the unskilled user. Another disadvantage is the relative difficultly of acquiring and matching DTDs and navigation files for SGML documents. The researcher obtained an SGML encoded version of all the books of the Old and New Testaments of the King James version of the Bible (OCLC Home Page) to test the formatting and search capabilities of Panorama. Navigation and stylesheet files were designed for the SGML Bible, while a TEI "rules" file (adapting the TEI DTD for use in Panorama) was obtained from CETH. The resulting electronic Bible was browsed and searched and yielded satisfactory retrievals from keyword proximity searches. A major weakness of the retrievals, however, was the absence of reference indicators (Book, chapter, verse) in the display, even though these were accurately marked up and could be viewed in the "Contents" frame.

Other applications such as Folio Views, MS Access, Q&A Database, RealPlayer etc. are discussed earlier in the thesis.

Glossary of Terms

The purpose of this chapter is to provide a glossary of terms used in the dissertation, but also a resource for learners of hypertext techniques. The items are arranged in alphabetical order in the print version of the dissertation, but may also be accessed under the following headings in the hypertext version: Hypertext, Digital Imaging, Internet, and Computers. Readers of the hypertext version of the dissertation will find that much of the glossary may be accessed via hyperlinks in the other chapters. Those with Internet access may access wider resources on the World Wide Web directly from links embedded in the text. Unless otherwise indicated, definitions and explanations are adapted from Besser and Trant (1995) and Dowining (1997).

ASCII (American Standard Code for Information Interchange) A standard code for representing characters as numbers that is used on most microcomputers, computer terminals, and printers. In addition to printable characters, the ASCII code includes control characters to indicate carriage return, backspace, etc. A popular extension of ASCII is the ANSI CHARACTER SET.

Bit-mapped image. An image created from a series of bits and bytes that form pixels. Each pixel can vary in colour or gray-scale value. (E.g. ".bmp" file format) Also, digital image.

Boolean Query A query formed by joining simpler queries with and or, and not. For example: "Find all books with author 'Dowining' and subject 'Computers' and not published before 1987." (Boolean algebra was developed by George Boole in the 1850s) See also Full text search.

Browser A computer program that enables the user to read hypertext on the World Wide Web. Popular World Wide Web browsers include Netscape and Microsoft Internet Explorer. Systems that lack graphics can use Lynx. (See World Wide Web; HTML)

CD-ROM Compact Disc Read-Only Memory. A form of write-once, disc-based, random-access data storage, usually mass-produced and distributed as a publication. At present, capable of holding approximately 550 megabytes of data. (One megabyte stores approximately 250,000 words of text. The King James Version of the Bible contains 820,270 words, so one compact disc could store slightly more than 550 full-text books the same length as the bible).

Client/server A systems architecture that divides functions (which might be part of a single application) between two or more computers. The client is the machine that requests the information; the server is the machine that supplies it. In the case of the *Encyclopaedia* database, the University of Durban-Westville server stores a copy of the database file listing authors, titles, etc. while the client (for instance a student at another university) may search the file using her internet browser.

Compression/Decompression Compression is the process of squeezing more data into a smaller storage space. Decompression is the retrieval of compressed data and its reassembly to resemble its original form (before compression).

Database A collection of data stored on a computer storage medium. such as a disk, that can be used for more than one purpose. For example, a firm that maintains a database containing information on its employees will be able to use the same data for payroll, personnel, and other purposes. See Database management.

Database management The task of storing data in a database and retrieving information from that data. There are three aspects of database management: entering data, modifying or updating data, and presenting output reports. Many mainframe computers are used by businesses for database management purposes. Several software packages are available for database management on microcomputers, such as dBASE Microsoft Access, and Symantec Q&A and some data management capabilities are provided with spreadsheets such as Lotus 1-2-3 and Excel. Some examples of database applications include maintaining parts order lists and keeping track of inventories; maintaining customer lists, and preparing bills for credit customers; and keeping track of the students at a school.

Information in a database system is generally stored in several different files. For example, a business will often have a file of regular customers and a file of employees. Each file consists of a series of records, such as one person or one transaction. Each record consists of several fields with each field containing an individual data item. For example, in an employee file there would be one record for each employee and there would be a field containing the person's name, a field for the address, a field the Social Security number. A database management system must make provisions by adding new records (e.g., when an employee is hired); for deleting unneeded records (e.g., when an employee retires); and for modifying existing records. Some fields (such as the Social Security number) will not change; other fields (such as year-to-date pay) must be changed frequently. The main purpose of a database management system is to make it possible to obtain meaningful information from the data contained in the database. A database program can respond to brief queries on the screen, or it can present detailed printed reports in a format chosen by the user. Here are some general functions that a database management system should be able to fulfill: (a) sort the records according to the order indicated by one specific field (e.g., sort in alphabetical order by name, or in numerical order by zip code). You should be able to designate a secondary field along which sorting will occur when there are ties in the primary field. For example, if you are sorting the records by the number of months the customers are overdue in their payments, you probably would like the names of all people 1 month overdue in

alphabetical order, then the names of all people 2 months overdue in alphabetical order and so on. (b) Set up selection criteria that allow you to examine only the records that meet a specific condition. For example, you may wish to look only at customers who live in your city, or you may wish to look at all employees whose job title is either "delivery driver" or "warehouse worker." (c) Count the number of records that meet a specific condition. For example, you may wish to count the number of employees who have been with the company for more than 10 years. (d) Perform calculations, such as computing the total amount owed on overdue accounts, or the year-to-date pay or each employee. (d) Connect information from more than one file. For example a database system might contain an employee tilt that lists the job classification for each employee. A separate file for each job classification would contain information on wages, fringe benefits, and work schedules that apply to all workers in that classification. (See also Relational Databases.)

DPI Dots per inch. A measurement of the scanning resolution of an image or the quality of an output device. Expresses the number of dots a printer can print per inch, or monitor can display, both horizontally and vertically. A 600-dpi printer can print 360,000 (600 x 600) dots on one square inch of paper.

Electronic document A document intended to be read as it is displayed on a monitor. Since it is freed from the constraints of printing, an electronic document can make use of hypertext, screen special effects such as flashing, animations, and full colour. Web pages are a type of electronic document; so are catalogues, documentation, and multimedia presentations distributed on CD-ROM. See also AUTHORING SOFTWARE.

Electronic mail (e-mail) the transmission of messages by computer from one person to another. Messages are saved until the recipient chooses to read them. Email is much more convenient than ordinary mail or telephone calls because it arrives immediately but does not require the recipient to be present, nor does it interrupt anything else the recipient may be

doing. Messages are easily printed out, saved on disk, or forwarded to other people. All users of email should be aware that backup copies of the messages may be saved on disk or tape and that perfect privacy cannot be guaranteed.

Electronic publishing (a) The creation, manufacturing, and distribution of paperless documents. Examples of "electronic documents" are CD-ROM encyclopaedias and the home pages on the World Wide Web. Each of these new formats brings new challenges and technical problems, but all to produce a pleasing combination of graphics and text. (b) The use of dedicated (devoted to one purpose), computer-controlled equipment in the publishing and printing industries. Desktop publishing may be considered part of this trend, but electronic publishing encompasses the use of specialised equipment not readily available to the mass market Electronic publishing is superseding traditional methods of pre-press production.

Front end This is a computer or a program that helps you communicate with another computer or program. For example, supercomputers usually do not communicate with their users directly; instead users submit programs through another computer called the front end. Similarly, an Internet browser acts a front end to information stored on other computers, and a hypertext system like Folio Views is a front end to text and multimedia files stored in other formats.

Full text search the act of searching through every word in a set of documents to retrieve information you are interested in. This is a slow but thorough way to use a computer to search through court records, scholarly journals or other material to retrieve items that you are interested in. An example of a full text search system is Lexis, which is available to lawyers by modem. It allows the user to search through a wide range of legal documents, such as court decisions and law review articles. The user specifies a search target, and the

system will find all documents meeting the specific condition. For example, if you are researching computer privacy, you can have the system find all documents that contain "computer" and also contain "privacy" within 25 words of the same place. This is better than searching for the specific phrase "computer privacy" (which might not occur in some documents on computer privacy) or searching for "computer" and "privacy" individually (each of which occurs in too many irrelevant documents). Search terms can also be connected with the word OR. This is most useful when the same item might be relocated to more than one way, such as "Holland OR Netherlands" or "Visa OR " See also Boolean Query.

Flatbed scanner An image-capturing device resembling a photocopy machine. The object to be scanned is placed face-down on a glass plate. The CCD array passes beneath the glass.

GIF Graphic Image File format. A widely-supported image-storage format promoted by Compuserve that gained early widespread use on on-line services and the Internet.

HTML Hypertext Markup Language An encoding format for identifying and linking electronic documents use to deliver information on the World Wide Web.

Hypertext (or hyperdocument) electronic documents that present information that can be read by following many different connections, instead of just sequentially like reading a book. Microsoft Windows help files and CD-ROM encyclopaedias are examples of hypertext. A hypertext document typically starts with it computer screen (full of information (text, graphics, and/or sound). The user then will have different options as to what related screen to go to next; typically, options are selected by a mouse. Encyclopaedia information is especially suitable for hypertext presentation. Each entry can be a screen of information, and each cross-reference can be a button that the user can click on in order to jump to that entry. Software help files are also a good application for hypertext because the user generally needs

to find a particular piece of information as quickly as possible rather than reading through the whole document. There is a danger that the user might become lost in the middle of a hyperdocument. A good hyperdocument should include some form of navigational aid that allows the user to see an overview of the document. Also, it is helpful if the computer maintains a record of the path that has been followed, both so the user can go backwards and so it is possible to retrace the same path at a future date if so desired.

Internet A cooperative message-forwarding system linking computer networks all over the world. Users of the Internet can exchange electronic mail, participate in electronic discussion forums (newsgroups), send files from any computer to any other via FTP, retrieve information via Gopher or HTTP and even use each other's computers directly via Telnet or logon if they have appropriate passwords. (Also World Wide Web; Information superhighway.)

Every user of every machine on the Internet has an address. For example, the address udw.ac.za:80/~stewartg/

means:

udw University of Durban-Westville

ac Educational site

za South Africa

80 Default server name

~stewartg Individual user sub-directory

Here udw.ac.za a domain address that gets translated into a numeric IP address, in this case 192.96.21.252, by the network itself. U.S. commercial, government, and military sites have addresses that end in com, gov, and mil respectively. Other countries have distinctive suffixes, such as.za for South Africa and .uk for Great Britain. The cost of running the Internet is paid largely by the sites that receive messages, and the sites that pass them along, not by the sites that send messages out. This has important legal and ethical implications.

Advertising on the Internet is almost always unwelcome, as is any self-serving misuse of electronic communications, because the sender of the material is not paying the cost of distributing it. The Internet grew out of the ARPAnet (a U.S. Defense Department experimental network) as well as BITNET, Usenet, and other wide area networks. Contrast INTRANET.

Intranet Theopposite of the Internet; a single network confined to a single organisation (but not necessarily a single site).

JPEG Joint Photographic Experts Group. Used to refer to the standard they developed for the still-image compression, which is sanctioned by the International Standards Organisation (ISO).

MPEG Motion Pictures Experts Group. Used to refer to an image-compression scheme for full-motion video they developed, which is ISO-sanctioned. MPEG takes advantage of the fact that full-motion video is made up of many successive frames, often consisting of large areas that don't change, such as blue sky background. MPEG "differencing" notes differences, or lack of them, form one frame to the next.

Navigation Finding your way through a complex system of menus or help files. This can be a real challenge, but there are a few tricks to help you. Programs for viewing hypertext files (such as help files and web pages) usually have a command called "Back" that allows you to backtrack to the previous screens. This is similar to Tom Sawyer using a rope to find his way around caves. A frequent frustration is to have a vague memory of a subject you read about yesterday, but can't remember how to get there. Familiarise yourself with the "Search" capabilities of the help file; it can save you a lot of time. As always a good index is worth its weight in gold. If the index is too general to be useful. written a complaint to the software vendor. (If enough users complain, something might be done.) In the meantime, you may

want to make a few notes on an index card and slip it into the manual. Learning how to navigate menus requires an adventurous spirit. Make yourself a map (if there's not one already in the manual), as any good explorer would do. Sometimes the logic of grouping certain commands together will not be apparent to you and you'll have to earn some rather arbitrary distinctions. The best defence is to be familiar with your software. If you know that there is a command to do whirligigs but can't remember whether it's under "File" or "Arrange": it's only a matter of a fraction of a second to look under both categories. Menus can nest like the wooden Russian dolls. One will lead to another in a rather infuriating way. Just remember to take one thing at a time. After making your decisions at each level, click "OK." If you've become lost in the menus, you can back out at any time by choosing "Cancel". Note: if you cancel out, the changes you made will not take place. Be aware that menus can interconnect at lower levels. This means that there can be more than one way into the same DIALOG BOX. Also "up" is not necessarily the same direction as "back".

Newsgroup An Internet discussion group devoted to a particular topic.

OPAC On-line Public Access Catalogue. A common term for automated, computerised library catalogues, made available to a wide range of users. (SABINET provides a gateway to South African library OPACs)

Object linking and embedding (OLE) (in Microsoft Windows 3.1 and later versions) a method of combining information that is processed by different application programs, such as inserting a drawing or a portion of a spreadsheet into a word processing document. The main document is called the client and the document or application that supplies the embedded material is the server. OLE supersedes an older feature of Windows called dynamic data exchange (DDE). OLE can be done either of two ways. An embedded object becomes part of the document that it is inserted into. For example, if you embed a drawing into a word processing document, the whole thing becomes one file, and to edit it, you use

the word processor, which will call up the drawing program when you double-click on the drawing to edit it. An embedded object has a life of its own; it remains a separate file and can be edited separately. When you edit it, the information that is linked from it into other documents is automatically updated. Thus, you can use a word processor to create a report that has links to a spreadsheet, and when you update the information in the spreadsheet, the corresponding information in the Import will be updated automatically Embedding and linking corresponding to "cold links" and "hot links" in Windows 3.0 DDE.

Optical character recognition (OCR) The recognition of printed or handwritten characters in an image of a piece of paper. OCR software is commonly used with scanners so that information received on paper will not have to be retyped into the computer. A difficulty is that the computer usually cannot recognize letters and digits with complete certainty, so it has to make intelligent guesses based on the spellings of known words. For trample, if you type "chack" an OCR device is likely to read it as "check." Obviously. OCR has difficulty distinguishing 1 (number one) from l (lowercase L) or 0 (number zero) from O; so do humans if they don't know the context. Information obtained through OCR should be carefully checked for accuracy. See also Flatbed scanner.

Pixel The picture elements that make up an image, similar to grains in a photograph or dots in a half-tone. Each pixel can represent a number of different shades or colours, depending upon how much storage space is allocated for it.

Resolution A measure of the amount of detail that can be shown in the images produced by a printer or screen. For instance, many laser printers have a resolution of 300 dots per inch (DPI), meaning that they print characters using a grid of black or white squares each 11300 of an inch across. This means that their resolution is 150 lines per inch printing line art, or 50 lines per inch printing halftone shadings (e.g. photographs), which use pixels in groups of 6. The human eye normally resolves about 150 lines per inch at normal reading distance, but a

person examining a page critically can distinguish two or three times this much detail. The resolution of a screen is given as the total number of pixels in each direction (e.g., 640 x 480 pixels across the whole screen). The equivalent number of dots per inch depends on the size of the screen. Present-day video screens resolve about 80 dots per inch; they are not nearly as sharp as ink on paper. A big advantage of draw programs, as opposed to paint programs, is that they can use the full resolution of the printer; they are not limited to printing what they display on the screen. However some paint programs can handle very detailed images by displaying only part of the image at a time.

Telnet A command that lets you use your computer as a terminal on another computer through a network; widely used on the Internet. Some versions are known as "tn" rather than telnet. Normally, the telnet program provides a direct path so that the remote computer communicates directly with the terminal you are actually using. See Internet.

Text file A file that contains lines of written information that can be sent directly to the screen or printer by using ordinary operating system commands. The files produced by word processors are usually not text files. Although they contain text, they also contain special codes underlining, etc.) whose meaning is known only to the word processing software. Many word processors can, however, produce text files ("text only" in MS Word, "DOS text files" in WordPerfect). On machines that use the ASCII character set, text files are often called ASCII files. To create a text file on the Macintosh, save a document as "text only."

TIFF Tagged Image/Interchange File Format. A file-storage format implemented on a wide array of computer systems. Considered an industry standard.

URL Uniform Resource Locator. A standard addressing scheme used to locate or reference files on the Internet. Used in World Wide Web documents to locate other files. A URL gives

the type of resource (scheme) being accessed (e.g., gopher, ftp) and the path to the file. The syntax used is scheme://host.domain[:port]/path/filename.

Video digitiser An image-capture device that employs a video camera (or video recorder) attached to a circuit board in a computer which converts the video signal into a digital file. Also called a frame-grabber.

World Wide Web WWW. An interconnected network of electronic hypermedia documents available on the Internet. WWW documents are marked up in Hypertext Markup Language (HTML). Cross references between documents are recorded in the form of URLs.

Word processing The process of using a computer to prepare written documents (letters, reports, books, and the like). The boundary between word processing and desktop publishing is not sharp but in general, word processing is the preparation of clearly worded, readable text, and does not include elaborate design or typography. Word processing makes it easy to change or correct a document and then print it out without introducing new errors.

References

1994. *The Oxford English Dictionary.* Second Edition. On Compact Disc. Oxford: Oxford University Press.

1996. *Discussion Document: Lifelong Learning through a National Qualifications Framework.* Committee for Development Work on the NQF.

1996. "The Victorian Women Writers Project". *The Source: the Newsletter of the Indiana University Libraries.* 1.2. Internet. URL http://www.indiana.edu/~libadmin/source/Fall96/victoria.html

1998. Murdoch University's Teaching and Learning Working Party Report. Internet. URL http://cleo.murdoch.edu.au/asu/ edtech/webtools/compare.html

1998. "White Paper: Distributed Learning". Lotus Institute Home Page. Internet. URL http://www.lotus.com/

Adey, D., R. Beeton, M. Chapman, and E. Pereira. 1986. *Companion to South African English Literature.* Craighall: Ad. Donker.

Andrews, D. and M. Greenhalgh. 1987. *Computing for on-scientific applications.* Leicester: Leicester University Press.

Angenot, M. 1989. *1889. Un état du discours social.* Longueil, Qué.: Préambule.

ARTFL. 1996. "ARTFL: A textual database. 2000 Texts 13th-20th Centuries *(Trésor de la Langue Française)*". ARTFL Project Home Page. Chicago: Centre National de la

Recherche Scientifique and The University of Chicago. Internet. URL
http://humanities.uchicago.edu/ARTFL/ARTFL.html

Barthes, R. 1977. "The death of the author". *Image, Music, Text: Roland Barthes* Ed. S.
Heath. New York: 146.

Bearman, D. 1995. "Toward a research agenda for cultural heritage on information networks
- Introduction". *Research Agenda for Cultural Heritage on Information Networks.*
Santa Monica: The Getty Art History Information Program. September, 1995.

Behrens, S.J. 1994. *Bibliographic control and information sources.* Pretoria: University of
South Africa.

Berners-Lee, T. and R. Cailliau. 1990. "World Wide Web: Proposal for a HyperText
project". CERN Website. Internet. URL http://www.w3.org/Proposal.html

Besser, H. and J. Trant. 1995. *Introduction to Imaging. Issues in Contructing an Image
Database.* Santa Monica: The Getty Art History Information Program.

Birkerts, S. 1997. "Introduction: The Reading Wars". *The Gutenberg Elegies.* New York:
Faber.

Boaz, J.K. and M. Boaz. 1996. "T.S. Eliot on a CD-ROM: A narrative of the production of
a CD". *Computers and the Humanities.* 30:1. 131-138.

Bradley, J. 1991. "TACT design". *CCH Working Papers.* Centre for Computing in the
Humanities, University of Toronto: 1.

Bradley, J., L. Presutti and M. Stairs. 1994. *TACT 2.1*. Computer Program. Centre for Computing in the Humanities, University of Toronto.

Brown University. 1996. "The Brown University Women Writers Project". The Brown University Women Writers Project Home Page. Providence: Brown University. Internet. URL http://www.wwp.brown.edu.

Bruce, D. 1993. "Towards the implementation of text and discourse theory in computer assisted textual analysis". *Computers and the Humanities*. 27: 357-364.

Busa, R. 1992. *Thomae Aquinatis Opera Omnia cum Hypertextibus*. Milan: EDITEL. CD-ROM.

Bush, V. 1945. "As we may think". *The Atlantic Monthly*. 176: 101-108.

Catano, J.V. 1979. "Poetry and computers: experimenting with the communal text". *Computers and the Humanities*. 13: 269-275.

Coover, R 1992. "The end of books". *New York Times Book Review*. 21 June.

Corns, T.N. 1991. "Computers in the humanities: methods and applications in the study of English Literature". *Literary and Linguistic Computing* 6.2: 127-130.

Cronje, J.C. 1998. "Interactive Internet: using the Internet to facilitate co-operative distance learning". *Proceedings: Weaving a Web of Learning*. University of Pretoria. March.

Cumming, A. and G. Sinclair. 1991. "Conceptualizing hypermedia curricula for literary studies in schools". *Hypermedia and Literary Studies*. Eds. G.P. Landow and P. Delaney. Cambridge, Massachusetts: The MIT Press. 315-328.

Cummings, A.M. et al. 1992. "Electronic publishing". *University libraries and scholarly communication*. Santa Monica: The Association of Research Libraries.

Derby, C. 1998. "Asynchronous Learning Networks." *Proceedings: Weaving a Web of Learning*. Pretoria: University of Pretoria, March.

Dowining, D.A. et al. 1997. *Dictionary of computer and internet terms*. New York: Barron's.

Elam, K. 1980. *The Semiotics of Theatre and Drama*. London: Methuen.

Foucault, M. 1972. *The archaeology of knowledge*. London: Routledge.

Fourie, I. and S.J. Behrens. 1997. "'South African Studies': a review of NISC's anthology CD-ROM". *South African Journal of Library and Information Science*. 65, 2: 116-123.

Garson, P. 1996. "Police slow to tackle piracy". *Mail & Guardian*. May 31 - June 6, 1996: 9.

Giordano, R. 1994. "Notes on operations. The documentation of electronic texts using text encoding initiative headers: an introduction". *Library Resources and Technical Services*. 38.4: 389-401.

Graves, J.K. 1997. "Design considerations for the Library of Congress Learning Page: Providing learners context and access to the collections". *Library Trends*. 45:4. 676-686.

Greenstein, D.I. 1996a. "Connecting Scholarly Communities and Networked Resources. The Arts and Humanities Data Service and the Urgency of Collaborative Endeavour". *Proceedings of the 1996 CAUSE Annual Conference*. 7-6.

Greenstein, D.I. 1996b. "A gentle introduction to historical data and databases." *Electronic Texts in the Humanities: Methods and Tools. General Handbook*. Princeton: Centre for Electronic Texts in the Humanities. Section 8: 1-29.

Hall, W. 1996. "Zen and the art of linking". Proceedings: Hypermedia '96. 3rd Southern African Conference on Multimedia and Hypermedia Pretoria: University of Pretoria, 22-24 May, 1996.

Hannabus, S. 1996. "Foucault's view of knowledge". *Aslib Proceedings*. 48.4: 87-102.

Harner, J.L. ed. 1996. *The World Shakespeare Bibliography. 1990-1993*. Cambridge: Cambridge University Press. CD-ROM.

Hart, M. 1992. "History and philosophy of Project Gutenberg." Project Gutenberg Home Page. Online. Internet. Available http://jg.cso.uiuc.edu/pg_home.html

Hart, M.S., P. Groves and G.D.J. Stewart. 1998. "Bookless Places". Internet. Discussion group ETEXTCTR-L@cornell.edu.

Hockey, S. 1994. "Electronic texts in the humanities: a coming of age". *Literary Texts in an Electronic Age*. Ed. B. Sutton. University of Illinois at Urbana-Chamapign: Graduate School of Library and Information Science.

Hockey, S. 1995. "Knowledge representation". *Research Agenda for Cultural Heritage on Information Networks*. Santa Monica: The Getty Art History Information Program. October, 1995.

Hoogcarspel, A. 1994. "The Rutgers Inventory of Machine-Readable Texts in the Humanities: cataloging and access". *Information Technology and Libraries*. March: 27-34.

Ide, N. and J. Véronis. 1995a. "Encoding dictionaries". *Computers and the Humanities*. 29: 167-179.

Ide, N. and J. Véronis. 1995b. "Introduction". The Text Encoding Initiative: Background and contexts. *Computers and the Humanities* 29. pp. 3-4.

Janus, R.J. 1997. "From paper and ink to CD-ROM: Digitizing the World Book image". *Library Trends*. 45:4. 602-622.

Jensen, M.B. 1992. "Making copyright work in electronic publishing models". *Serials Review*. Spring & Summer, 1992: 62-65.

Kahn, P. 1991. "Linking together books: experiments in adapting published material into intermedia documents". *Hypermedia and Literary Studies*. Eds. G.P. Landow and P. Delaney. Cambridge, Massachusetts: The MIT Press. 221-256.

Katchoff, B. and M. Ryan. 1997. "Computer Mediated Communication (CMC) - a tool for Open and Distance Learning (ODL)". Unpublished paper.

Klopper, D. 1990. "Politics of the Pastoral: The poetry of Thomas Pringle". *English in Africa*. 17:1. 21-59.

Kolb, D. 1994. "Socrates in the labyrinth". *Hyper/Text/Theory*. Ed. G.P. Landow. Baltimore and London: The Johns Hopkins University Press. 323-344.

Landow, G.P. 1989. "Hypertext in literary education, criticism, and scholarship". *Computers and the Humanities*. 23: 173-198.

Landow, G.P. 1994. "What's a critic to do?: Critical theory in the age of hypertext". *Hyper/Text/Theory*. Ed. G.P. Landow. Baltimore and London: The Johns Hopkins University Press. 1-48.

Landow, G.P. and P. Delany. 1991. "Hypertext, hypermedia and literary studies: the state of the art". *Hypermedia and Literary Studies*. Eds. G.P. Landow and P. Delaney. Cambridge, Massachusetts: The MIT Press.

Lewis, T. M. Gould. and M. Ryan. 1997. "Computer Conferencing in the Continuing Professional Development of Teachers in the Post-16 Sector - a Case Study". *Electronic Pathways: Adult learning and the new communication technologies*. Ed. J. Field. NIACE.

Lyotard, J-F. 1979. *The postmodern condition: a report on knowledge*. Manchester: Manchester University Press.

Mah, C. and J. Flanders. 1996. "Scholarly needs, encoding challenges: correction, regularization, and expansion". *The Brown University Women Writers Project Newletter*. 2.2. 1-7. Summer.

Marchionini, G. 1995. "Resource search and recovery". *Research Agenda for Cultural Heritage on Information Networks*. Santa Monica: The Getty Art History Information Program. October, 1995.

McCarty, W. 1996. "Introduction to concording and text-analysis: history, theory and methodology". *Electronic Texts in the Humanities: Methods and Tools. General Handbook*. Princeton: Centre for Electronic Texts in the Humanities. Section 5: 1-17.

McCollum, K. 1997. "Colleges sort through vast store of tools for designing web courses." *The Chronicle of Higher Education: Information Technology*. 21 October, 1997.

Mda, Z. 1995. *The Hill*. In *Open Space. Six Contemporary Plays from Africa*. Eds. Y. Hutchinson and K. Omotoso. Cape Town: Kagiso Publishers.

Mills, J. and B. Chandramohan. 1996. "Literary studies: a computer assisted teaching methodology". *Computers and the Humanities*. 30:1. 165-170.

Nelson, T.H. 1989. "Hyperwelcome". *Hypermedia 1*. 1: 3-5.

Nelson, T.H. [1965] 1987. *Literary Machines. Edition 87.1. The report on, and of, Project Xanadu concerning word processing*. Swarthmore, PA: Theodor H. Nelson.

Nielsen, J. 1995. *Multimedia and Hypertext: The Internet and Beyond*. Boston: AP Professional.

NISC 1996. *South African Studies CD-ROM*. Grahamstown: National Inquiry Services Centre.

Olsen, M. 1993. "Signs, symbols and discourses: a new direction for computer-aided literature studies". *Computers and the Humanities*. 27: 309-314.

O'Prey, P. Ed. 1983. J. Conrad. *Heart of Darkness*. New York: Penguin. Introduction.

Papert, S. 1986. "Logo as a Trojan Horse: Reflecting Logo philosophy in the context of real school experience". *Logo86 Conference Pre-Proceedings*. Cambridge, Mass.:MIT.

Potter, R. 1996. "What computers are good for in the literature classroom". *Computers and the Humanities*. 30:1. 181-190.

Pringle, T. 1837. *The Poetical Works of Thomas Pringle* Ed. L. Ritchie. London: Edward Moxon.

Pringle, T. [1834] 1966. *Narrative of a Residence in South Africa*. Ed. A.M. Lewin Robinson. Cape Town: C. Struik, Africana Collectanea Vol. XX.

Robinson, P.M.W. 1996. "Some electronic editions: Johnson's Dictionary, the World Shakespeare Bibliography and the Canterbury Tales". *Electronic Texts in the Humanities: Methods and Tools. General Handbook*. Princeton: Centre for Electronic Texts in the Humanities. Section 7: 1-24.

Rockwell, G. 1996. "Hypertext in the humanities". *Electronic Texts in the Humanities: Methods and Tools. General Handbook*. Princeton: Centre for Electronic Texts in the Humanities. Section 9: 1-19.

Ryan, M. and G. Culwick. 1997. "Creating a virtual integrated learning environment (VILE)". *ED-MEDIA and ED-TELECOM 97 Proceedings*. Eds. T. Muldner, and T. Reeves. Calgary, Canada, June 1997: AACE, Virginia.

Seaman, D.M. 1994. "'A library and apparatus of every kind': The Electronic Text Center at the University of Virginia". *Information Technology and Libraries*. March, 1994: 15-19.

Senge, P. et al. 1994. *The Fifth Discipline Fieldbook. Strategies and tools for building a learning organization*. London: Nicholas Brealey Publishing.

Serpieri, A. 1978. "On Rhetoric in the Theater". *Strumenti Critici*. 41.

Smit, J.A., A.J. van Wyk, and J-P, Wade. Eds. 1996a. *Rethinking South African Literary History*. Durban: Y Press.

Smith, K.W. 1997. "OCLC - Moving toward the next stage of the electronic library". OCLC Website. Internet. URL http://www.oclc.org/oclc/man/9680rldc/kwsmith.htm

Snyman, M.M.M. 1997. *Investigation into a future machine-readable cataloguing (MARC) format for South African Libraries: A Report with Recommendations*. Interim Committee for Bibliographic Organisation (ICBO).

Sperberg- McQueen, C.M. 1995. "Bare Bones TEI: A very small subset of the TEI Encoding Scheme". Document No. TEI U6. Chicago: Unversity of Illinois at Chicago.

Sperberg- McQueen, C.M. and L. Burnard. 1994. *Guidelines for electronic text encoding and interchange*. TEI P3. Chicago and Oxford.

Stewart, G.D.J. 1998. "Sharing South African Literature in Electronic Form: The Electronic Text Project at the Centre for the Study of Southern African Literature and Languages (CSSALL)". Abstract. Proceedings of the workshop on Data Sharing for Research Capacity Development. Pretoria: South African Data Archive, HSRC. 89-90.

Trenner, L. 1987. "How to win friends and influence people: definitions of user-friendliness in interactive computer systems". *Journal of Information Science*. 13: 99-107.

Van der Merwe, D. 1996. "The copyright nightmare with regard to multimedia". Proceedings: Hypermedia '96. 3rd Southern African Conference on Multimedia and Hypermedia Pretoria: University of Pretoria, 22-24 May.

Van Wyk, A.J. 1996a. "Towards a South African Literary History". In J.A. Smit, A.J. van Wyk, and J-P. Wade. Eds. *Rethinking South African Literary History*. Durban: Y Press. 31-39.

Van Wyk, A.J. 1996b. *Concise historical survey: South African Literature*. Durban: University of Durban-Westville.

Van Wyk, A.J. and G.D.J. Stewart. 1997. "The Shamanic Formation". Multimedia Presentation. In *SALIT Web* CD-ROM. Beta 01 Version. Durban: CSSALL.

Van Wyk, A.J., P. Conradie, and N. Constandaras. 1988. *SA in Poësie/SA in Poetry*. Pinetown: Owen Burgess.

Voss, A.E. 1990. "'The slaves must be heard;' Thomas Pringle and the dialogue of South African servitude". *English in Africa*. 17:1. May, 1990: 61-81.

Welsch, E.K. 1992. "Hypertext, hypermedia, and the humanities". *Library Trends*. 40.4: 614-646.

Whalley, B., G. Munroe et al. 1996. "Publishing a scholarly journal on the World Wide Web". *Aslib Proceedings*. 48.7/8: 171-176.

Wilson, B.G. 1997. "Thoughts on Theory in Educational Technology". *Educational Technology*. Jan/Feb: 22-27.

Wolff, M. 1994. "Poststructuralism and the ARTFL database: some theoretical considerations". *Information Technology and Libraries*. March, 1994: 35-42.

Zweig, R.W. 1997. "From newsprint to screen: lessons from the Palestine Post project". Unpublished article. Tel Aviv University.