Chapter 1

Introduction

Documents are not independent. Like biological organisms, every document is always related to some other (Brown and Duguid, 1996).

1.1 Motivation

The World-Wide Web is the most successful and farthest-reaching hypertext system to date, embodying a hypertext corpus (or *docuverse*) of enormous size which still continues to grow at a phenomenal rate. Researchers release their results to the Web before they appear in print; corporations list their Web URLs alongside their toll-free numbers; news media and entertainment companies vie for the attention of a worldwide browsing audience. The Web has come to reflect human culture — a massive social network encoding *links* among almost 10^9 documents (Pennock et al., 2002).

In hypertext theory, links are the mechanism by which hypertext is implemented. Hypertext began as a vision of the complete integration of texts (Bush, 1945), and has since become the vision of the complete integration of all information in any media (including text, image, audio, and video). The depth and diversity of the Web has tremendous potential to provide such an integrated docuverse. This thesis therefore considers the process of authoring and publishing *integrated* hypertexts in the Web docuverse — hypertexts which demonstrate not only *local coherence* (Moulthrop, 1992) (internal links between nodes), but also *global coherence* (links connecting the new contributions to the wider context of existing material published in the Web) — this thesis introduces the term *Associative Writing* to describe the creation of such integrated hypertexts.

From a user perspective, motivation for this work comes from the potential of integrated hypertexts to allow writers to demonstrate (using hypertext links) the innovation and significance of new ideas, show the reliability of the conceptual foundation being built on, and juxtapose existing ideas in previously unthought-of combinations. Further, writers

can express networks of facts and relationships by building corresponding networks of statements and links across the docuverse (VanLehn, 1985), extending the range of what can readily be said. Writers can link pithy ideas to an existing context, rather than having to use hundreds or thousands of words to establish a context, making even single-sentence publications useful and practical (Drexler, 1991). Links, in particular, will enable critics to attach their remarks to their targets, making criticism more effective by letting readers see it (Drexler, 1991). Readers not only benefit from the information they read in integrated hypertexts, but also from the richness of associations created by the writer; freedom of access within an hypertext structure provides a rich environment for understanding the information they find (Thüring et al., 1995).

A solid foundation for integrated writing in the Web could also have wider social implications. Berners-Lee's vision of the Web as "anything being potentially connected with anything" is a vision that "provides us with new freedom, and allows us to grow faster than we ever could when we were fettered by the hierarchical classification systems into which we bound ourselves...it brings the workings of society closer to the workings of our minds" (Berners-Lee, 1999). Drexler asserts that "knowledge is a primary asset of our civilisation, crucial to all our goals. By improving the quality of debate and speeding the evolution of knowledge, hypertext publishing will not only further our goals, but help us choose them more wisely" (Drexler, 1991).

1.2 Problem

Although evidence of Associative Writing strategies is evident in the Web (Carr et al., 2000b,a; Miles-Board et al., 2001), providing computer-based support for Associative Writing in the World-Wide Web docuverse is challenging on a number of different levels. This thesis reports on investigative work carried out in defining and overcoming five core challenges, and on practical work carried out in response to these challenges. The five challenges are outlined briefly below:

- The "lost in hyperspace" problem refers to "the tendency to lose one's sense of location and direction in a nonlinear document" (Conklin, 1987), and may be particularly applicable in integrated hypertexts in the Web.
- Legal issues surrounding "deep linking" to existing content published on the Web could prevent writers integrating new contributions with it.
- Links, the building blocks for integrated hypertexts, are restricted by the Web's hypertext model to uni-directional, binary links, which must be embedded in documents 'owned' by the writer: as a result the Web as a publication medium limits the expressive capabilities of the writer (reducing the range of what can 'be said').

- The Web's chaotic nature (documents are frequently edited, moved, or deleted in an ad-hoc manner) causes links to 'break', potentially permanently disconnecting integrated hypertexts from the existing context on which they build.
- Popular tools for writing new Web hypertexts do not adequately support the different cognitive activities involved in Associative Writing.

1.3 Approach

In response to the challenges outlined above, this thesis introduces the Associative Writing Framework (AWF), a framework which builds on existing open hypertext, Semantic Web, hypertext writing, and hypertext annotation approaches to provide a novel Web browser-based interface for supporting browsing/reading, annotation, link creation, and integrated writing.

Figure 1.1 illustrates AWF's innovative approach in a nutshell: the writer highlights interesting and relevant ideas in existing Web documents while browsing and reading using an integrated annotation component (in the form of a bespoke Internet Explorer toolbar); relationships between existing ideas are captured by 'drawing' links *directly on the screen* — these links can be assigned semantic types (note labelled arrows connecting ideas), and remain visible, following the movement of windows and scrollbars on the screen. As the writer's new contributions take shape (top left of figure), they build on the existing ideas and link structures uncovered and captured using the annotation and link drawing components. When the new contributions are published in the Web, readers can explore the global context of the writer's work, and may go on to build upon and challenge its interpretations in their own integrated writings.

A 'framework evaluation' has been successfully carried out as part of a case study of Associative Writing in the interpretation of dance performances, using AWF to capture intertextual relationships between Web-based media (Figure 1.2) and present them in the form of an integrated hypertext essay.

1.4 Thesis Structure

Chapter 2, *Hypertext and Associative Writing*, opens with a brief historical overview of hypertext systems, from the pioneering work of early visionaries to more recent work in open hypertext. The concept of Associative Writing is then discussed in more detail in the context of this historical perspective. A review of the meteoric rise of the World-Wide Web introduces the first two of several challenges — the controversial "lost in hyperspace" problem, and the recent copyright infringement claims arising from "deep linking" in the Web — the implications of which are discussed.

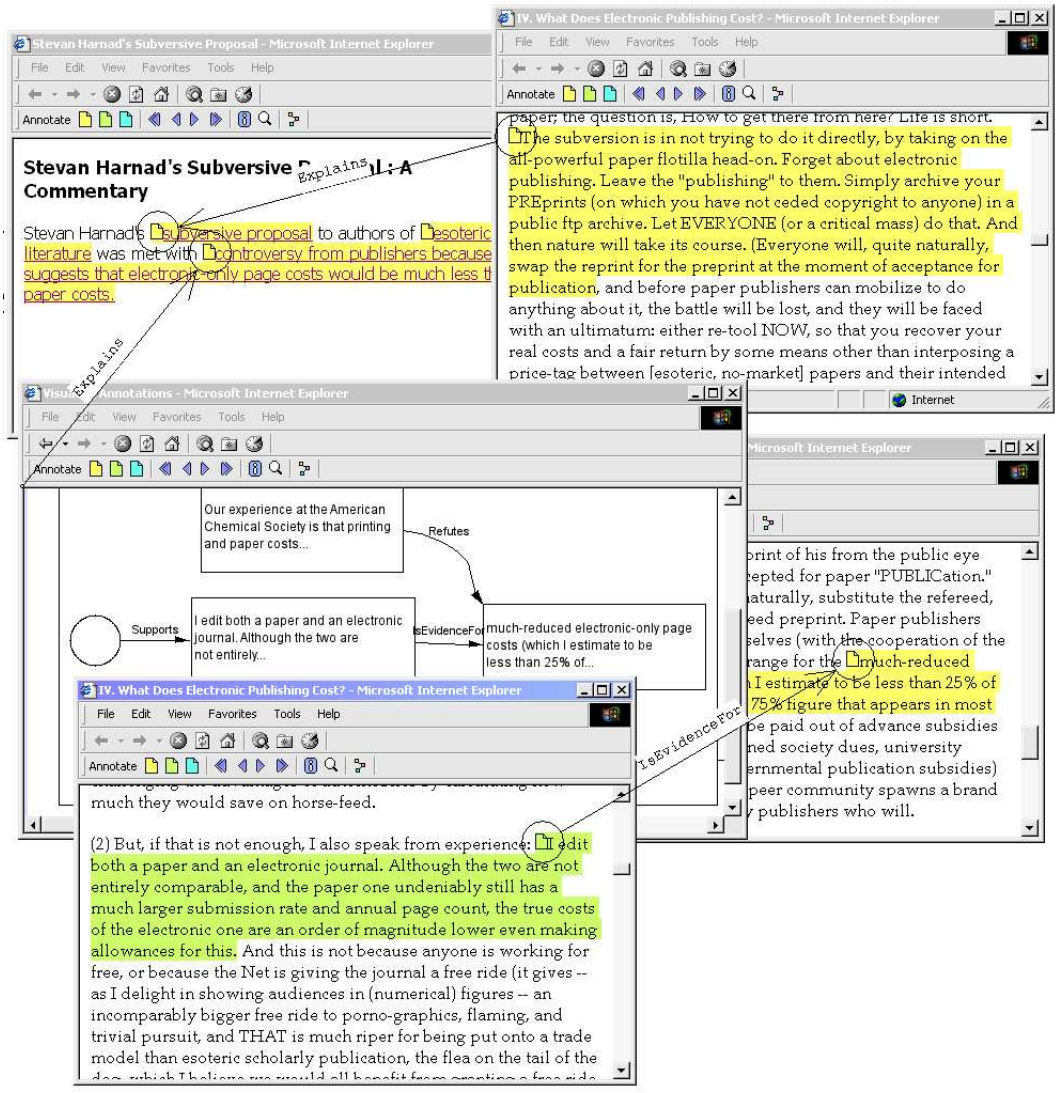

FIGURE 1.1: Integrated browsing, annotating, linking, and writing in AWF.

Chapter 3, *Searching for Associative Writing in the Web*, describes two investigations which aimed to uncover examples of Associative Writing in the Web for the purposes of illustration and discussion. Although the first (manual) investigation uncovered some significant examples, the number of examples was far less then anticipated, so a follow-up investigation conducted a systematic search for evidence of Associative Writing in the *Internet Archive*. The results of both investigations are presented and analysed, and also positioned relative to existing studies of Web linking phenomena which help inform proposals for future investigation.

Chapter 4, *Publishing Integrated Hypertexts in the Web*, explains how the Web's hypertext model limits the expressive capabilities of the writer, and ultimately the extent to which new contributions can be integrated with the docuverse. These limitations lead to the introduction of two further challenges — improving support for navigation and hyperstructures in the Web, and maintaining link integrity in the Web — and

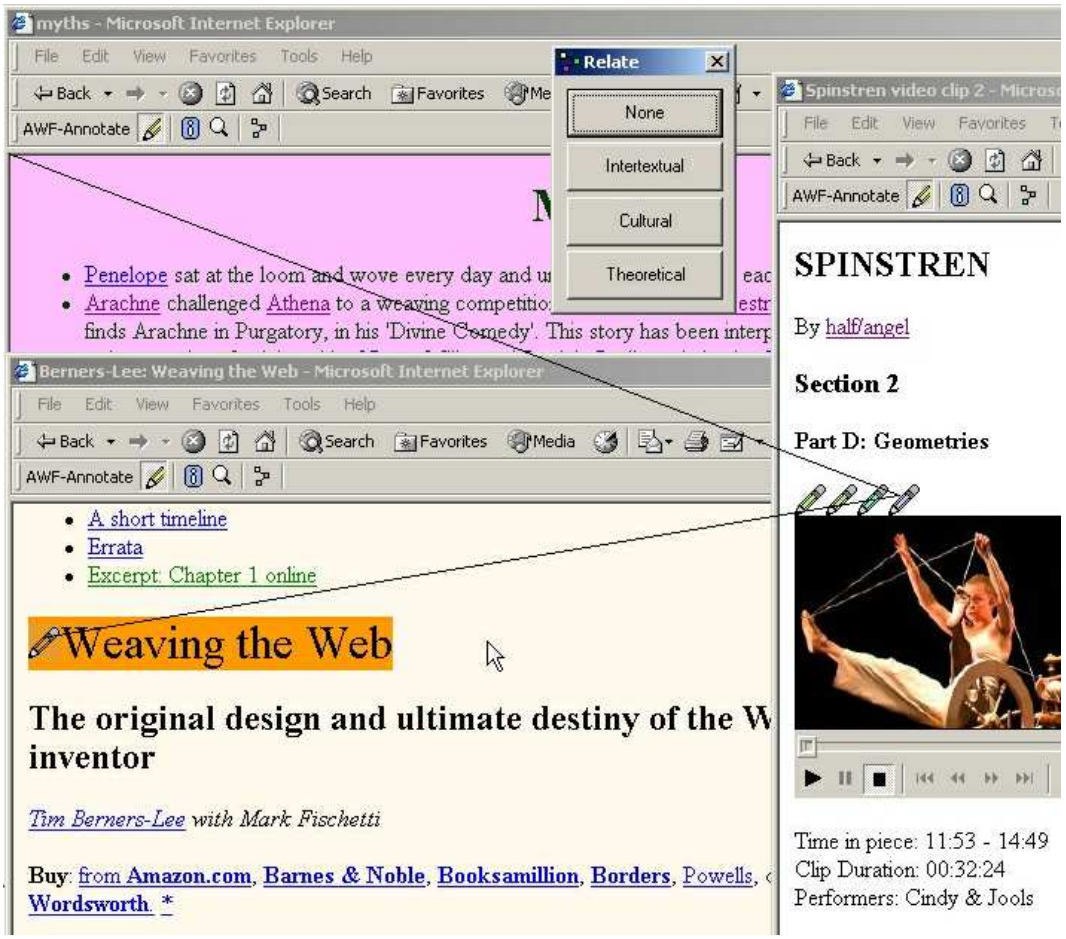

FIGURE 1.2: Capturing dance performance intertextualities with AWF.

the discussion of various approaches to these challenges using open hypertext and link management strategies.

Chapter 5, *Associative Writing and the Semantic Web*, investigates efforts to evolve the World-Wide Web into a machine-understandable Semantic Web, in an attempt to position this work relative to the current state of Semantic Web research, and also to demonstrate a potential convergence between supporting *human* understanding through Associative Writing, and *machine* understanding through Semantic Web technology.

Chapter 6, *Tools for Associative Writing*, introduces the final challenge addressed by this work: supporting the writer in carrying out the various activities involved in Associative Writing. The chapter starts by deriving an *a priori* model of Associative Writing from the correlations observed between a number of established hypertext writing models, and then uses this model to critically review the potential contributions made by a number of existing hypertext tools designed specifically to support writing tasks. The chapter concludes that the challenge should be focused on those Associative Writing activities which are least well supported by existing approaches.

Chapter 7, *Annotation: Integrating the Writer*, explores this focus in more depth by

investigating the possibility of leveraging existing hypertext, Web, and Semantic Web annotation technologies to support integrated writing.

Chapter 8, *A Framework for Associative Writing in the Web*, introduces the Associative Writing Framework (AWF), which has been specifically designed and implemented in response to the challenges discussed over the course of the preceding chapters. AWF unifies open hypermedia, link integrity management, Semantic Web, hypertext writing, and annotation approaches. The chapter begins by discussing the design of the AWF approach, and then describes the implementation of the framework using a simple (Web-based) integrated writing scenario to drive a tour of the framework features. A more detailed description of the software implementation is also provided before discussing how AWF responds to each challenge.

Chapter 9, *Case Study: Using AWF to Decentre the Dancing Text*, presents a case study of AWF as a tool for the analysis of dance performances; a focused, real-world scenario that demonstrates the validity of the Associative Writing model by analysing the process of dance analysis, and showing how AWF succeeds in meeting the requirements of a hypothetical 'dance hypertext' system. The chapter also reports on the positive feedback obtained from a 'framework evaluation' in this context, and concludes with a discussion of proposed future collaborations with University of Surrey dance experts.

Chapter 10, *Comparison of AWF with Previous Work*, compares and contrasts AWF with the work in open hypermedia, hypertext writing, annotation, and link integrity management on which the framework builds. A "unified framework" for the integration of AWF with existing approaches is also put forward.

Chapter 11, *Conclusions and Further Work*, concludes the thesis with a 'road-map' for proposed short- and longer-term future work directions.

1.5 Contributions

The following are presented as the core contributions of this work:

1.5.1 Web Analysis Data

In reporting the analysis of Web documents archived by the *Internet Archive* digital library, this work presents new data to the hypertext and Web communities which classifies and quantifies linking practices in the Web.

1.5.2 Associative Writing in the Web

This work has described a core set of challenges facing developers of tools or methodologies for Associative Writing in the Web, culminating in the Associative Writing Framework, a framework specifically designed in response these challenges.

1.5.3 Initial Evaluation

The ideas embodied by AWF have been evaluated in the specific real-world problem domain of dance analysis, in which a need for support in creating integrated hypertexts has been identified. Initial indications are that the framework method is valid, and that continued work is worthwhile.