14 Jan 1996

Paper read at the conference “Contemporary computer and network technologies. Unlimited possibilities to expand world-wide cultural and scientific bridges”, Moscow, Russia, 17-18 Jan. 1996

Internet and the humanities: the promises of Integrated Open Hypermedia

Donald Broady and Hasse Haitto
Royal Institute of Technology
Dept. of Numerical Analysis and Computing Science/CID
S-100 44 Stockholm
Sweden
E-mail `broady@nada.kth.se`, `haitto@nada.kth.se`
URL `http://www.nada.kth.se/~broady/`

Abstract

This paper explores some new possibilities opened by the development of IOH (Integrated Open Hypermedia) based on international standards, in particular SGML (Standard Generalized Markup Language, ISO 8879:1986) and its accompanying standards HyTime (Hypermedia/Time-based Structuring Language, ISO 10744:1992) and DSSSL (Document Style Semantics and Specification Language, ISO/IEC 10179:1996). SGML offers the opportunity to create common principles for structuring, encoding and exchanging material. HyTime offers the opportunity to create common principles for addressing and hypertext linking, as well as principles for structuring multimedia information (more precisely: time-dependent information as audio or video). DSSSL offers the opportunity to create common principles for the processing of SGML-encoded material (i.a., principles for portable style sheets). The most significant implementation of SGML in the humanities community is the comprehensive SGML-compliant guidelines for markup and exchange developed by the international project TEI (Text Encoding Initiative). The first half of the paper is more introductory in scope, the second half more technical.

1. Introduction: digital literature still in its infancy

Within the humanities there is a need for sophisticated tools enabling our capability to describe and capture innumerable kinds of *relations* within and between bulks of digital information.

This need is not always obvious. If all you wish to do is to publish fragments of historical sources, or a poem or a novel on a CD-ROM or on the World Wide Web for entertainment or similar purposes, your main concern might be to format and encode the material in order to deliver a pretty presentation on screen or on paper.

Many scholars wish to deliver more. They wish to present all kinds of relations: from the source to their own and others' commentaries, comparisons between different editions, comparisons between a close diplomatic representation of the author's manuscript and different versions resulting from the specialists' interpretations (a typical example is relations between abbreviations in mediaeval manuscripts and the proposed expansions of these abbreviations that you normally find in printed editions), relations to other primary sources, etc.

For paper publications there are many tools of this kind in existence, laboriously developed during the centuries and well known by philologists, historians etc. There are thus "de facto standards"—which may be competing with each other, but nevertheless discernible and functioning—prescribing how to establish a critical edition, how to create concordances and commentaries, how to cross-reference, how to point to a certain line on certain page in a book, even how to point to a certain paragraph or verse in a work published in various editions during the centuries, each with its own pagination and page layout.

This kind of capability is also to a large extent spread among the general public. Even small children know where to seek in a book for the title and the author's name. Many readers know how to use tables of contents, indexes, and footnotes. They are capable of finding their way through the hierarchical divisions into parts, chapters, sections, subsections, and paragraphs. Most of them are probably not aware that all these presentation technologies and many more are historical inventions of the last millennium, and some in fact of rather modern date. They would, though, realise this fact if they opened an ordinary medieval book. Even if they knew how to decipher the handwriting, the abbreviations and the orthographic extravagances the pages would appear as almost impenetrable homogenous patterns of writing. There is probably no title or author's name or other "front matter" in the beginning, no clear visual divisions in sections, no pagination, not even white spaces indicating divisions between individual words.

The late nineteenth century man's perception of a book originated in the scriptorium of some mediaeval monastery is probably rather similar to how people in a not too distant future will look at the electronic books and journals and other products from today's information providers.

This is not surprising. In fact, it sometimes takes a skilled eye to distinguish one of the first printed books from those written by hand. For centuries the printing houses aimed at results imitating the most able scribes. The same goes for today's presentation of digital literature. We still lack most of the technology, tools and apprehension abilities that future generations of readers will find indispensable in order to grasp the structure and content of digital literature.

2. Text Encoding Initiative

The Text Encoding Initiative has for more than six years engaged many hundred scholars in the humanities and computer scientists in many countries. TEI has been sponsored by the Association for Computers and the Humanities, the Association for Computational Linguistics, and the Association for Literary and Linguistic Computing, as well as by U.S. National Endowment for the Humanities, Directorate General XIII of the Commission of the European Communities, the Andrew W. Mellon Foundation, and the Social Science and Humanities Research Council of Canada.

The aim of TEI is to provide the humanities community with proposals for platform-independent encoding and exchange of a broad range of (mainly but not exclusively textual) information: transcriptions of ancient manuscripts or contemporary speech, poems, critical editions, scientific articles, etc. Numerous draft versions of TEI's guidelines have circulated on the internet, and a set of recommendations, called ”TEI P3” (that is the third set of propositions) was published in May 1994, edited by Lou Burnard at Oxford University and C.M. Sperberg-McQueen at Chicago University.

The P3 guidelines, ca. 1 300 pages in print, might be downloaded from, i.a., the URL http://etext.virginia.edu/TEI.html. More information on TEI and related efforts can be obtained by visiting for example http://www.sil.org/sgml/acadapps.html#ota. A TEI homepage hase recently been announced, URL http://www-tei.uic.edu/orgs/tei.

With the guidelines TEI is proposing not only encoding schemes but rather a way of organising information in order to make it useful for many purposes: print, screen presentation, searching, information retrieval, updating, exchange—all this in a portable and platform-independent format that bridges the gulfs between computer platforms, applications, national languages. The guidelines also represent an attempt to overcome the gulfs between competing scientific traditions and schools—some scholars prefer for example a critical edition that follows the primary sources extremely thoroughly, others might prefer a critical edition based on an extensive amount of interpretations and other editorial interventions. Some might be interested in the linguistic information, others in analyses of metaphors or themes or stylistics. Yet others might wish to produce an output to edit a non-scholarly edition, maybe a children's book version of a classic text. TEI is offering a modular and flexible approach. It is up to the user to tailor his or her own markup scheme.

The TEI guidelines are based on SGML, in a most innovative way that has had effects also outside the humanities community. The military organisations and large corporations (the automobile, aircraft, computer, pharmaceutics industry, etc.) have been adopting SGML much earlier and in a much larger scale than the scholars in the humanities, but today TEI has some impact in the military and business world.

2.1 Simple introduction to SGML markup

Those who are not familiar with SGML might need a simple hint on what this standard is all about [the following is a translation of a passage in Broady, 1995]

SGML (Standard Generalized Markup Language) was accepted as an international standard language for descriptive markup in 1986. It is not a markup scheme that prescribes what kind of markup (or “tags”) to put in your information. It is general language offering an international agreement on how to tell someone else (a human being or a computer system) how you have chosen to markup your information.

SGML is concerned with descriptive as opposed to procedural markup. Procedural markup serves the purpose of governing procedures, for example to tell a printer what to do. If you within a word processor apply so called “direct formatting” on a text passage in order to produce 12 points Courier italics on your printer, then you are performing procedural markup. In this case you are interested in the presentation of the information. When using descriptive markup your primary interest is in the structure and content of the information: what it is, not how it looks.

Take for example the following piece of text. There you find four occurrences of italics. They all look the same but they represent four types of elements: heading level 3, emphasis, gloss from a foreign language word, reference to a source.

> *1.3.1. On Descartes’ method*
>
> Descartes created a model for *the* philosophic method. Descartes’ celebrated text on the method was, however, published as a *préface* to three studies on natural sciences and mathematics (available in René Descartes: *Discours de la méthode & Essais*. Éd. Ch. Adam & P. Tannery. OEuvres VI. Vrin, Paris, nouvelle édition 1982).

Here is the same text provided with descriptive markup (using an early version of the tag sets proposed by TEI):

> **<h3>** On Descartes’ method**</h3>**
> **<p>**Descartes created a model for ****the**** philosophic method. Descartes’ celebrated text on the method was, however, published as a **<gloss><foreign lang=Fr>**préface**</foreign></gloss>** to three studies on natural sciences and mathematics (available in **<xref RID=DESCARTES1982>**).

The heading level 3 is enclosed by the starttag **<h3>** and the endtag **</h3>.** There is no need for a number if the system is able to count the chapters and sections and sub-sections and produce the correct number in the output. A paragraph is opened by **<p>** and closed by **</p>.** The emphasis is enclosed by the tags **** and **,** and the French word by the **<gloss>** and **</gloss>**, within which we find another pair of tags informing us (or the system, for example for spell checking) that this is a foreign word, and that the language is French. Finally the reference to another book is represented by a cross reference element with an ID attribute which makes it possible to retrieve the correct reference somewhere else in the document, perhaps in a reference list at the end.

In this case descriptive markup can help us to avoid that the same typographic look (italics) designates four different types of content elements. The other way around descriptive markup might help to avoid that one and the same element type is represented by several different looks. Quotations is one example. In different languages (and within each language) there are various typographical conventions when it comes to present inline quotations on paper or on screen. For example:

German	»XXXXX« ›XXXXX‹ „XXXXX“	French	« XXXXX »
		Swedish	”XXXXX” »XXXXX»
English	“XXXXX” ‘XXXXX’		

When applying descriptive markup you might not need to bother about these differences when editing a text. It might be sufficient to encapsulate the quotation within a pair of tags, **`<InlineQuote>`** and **`</InlineQuote>`**. The output system can use the appropriate typographic conventions in the output on paper or screen.

It is important to underline that the SGML standard does not prescribe what specific tags to use. It prescribes a way of defining, i.a., tag sets and their use, and thereby ways of describing the building blocks (elements, attributes) of a specific category of documents. Each category is defined by its own document type definition (DTD), used to tell someone (a human being or a system) which these building blocks are in this specific type of document, and the relations between the building blocks. You can define the parts of a sonnet, the various attributes to elements like person names, etc. The TEI guidelines thus offers modular ways of creating customised DTDs for almost every imaginable type of literature of relevance to the humanities. These DTDs are then used as a most powerful ingredients in SGML-aware systems. You can perform structure searches, as a search for all occurrences of a certain string within a certain pair of tags: search for all numbers lower than 1918 between the tags **`<PublicationDate>`** and **`</PublicationDate>`**. You can easily on the fly build tables of contents by extracting the headings from a document, as well as all kinds of indexes by extracting other elements (graphics, personal names, geographic locations, editorial comments). You can build style sheets in order to present different elements and attributes in different ways on paper or on screen, hiding some, highlighting or colouring others, thereby providing many different views of one and the same material. You can use one and the same SGML encoded “master document” to produce output in many different formats and on many different media: as word processing files, as HTML, as Braille, as PDF files, as PostScript files etc., for printing, screen presentation, on line delivery...

The possibilities mentioned are just a few examples of fairly simple uses of SGML, implemented in many widespread commercially available SGML software products. Much more sophisticated use of SGML is commonplace in many organisations and companies, and increasingly so also in the humanities. There are electronic libraries on their way, as well as large publication projects, extensive corpora to be used in literary or linguistic studies and many other implementations of SGML. Visit for example the website http://www.sil.org/sgml/sgml.html for more information on ongoing SGML software development and application projects.

3. World Wide Web and the heritage from the classic hypertext systems

The astonishing success of the World Wide Web (WWW) and the markup language HTML means that millions of users have today been exposed to some SGML markup principles, as well as with the heritage from the development of hypertext systems.

The scientific tradition of arguing and referencing previous work meshes well with the linking technology of the WWW. Indeed, the success of the WWW has convincingly demonstrated the appeal of globally interlinked data, of the docuverse envisioned by Vannevar Bush, Douglas Engelbart, Ted Nelson, and other hypertext pioneers.

However, the predominant document format on the WWW is HTML. The TEI, though geared primarily for documents marked up in SGML, provides not only methods of arbitrarily extending the tag set, but can also be used for other notations or data, where the TEI concept of a document header—information about the electronic file, or metadata as it is sometimes called—works as a wrapper, as an envelope, around the data.

The ISO standard HyTime (Hypermedia/Time-based Structuring Language ISO/IEC 10744:1992), an application of SGML, applies this idea in the form of a hub document to address virtually any kind of document and format, such as a movie clip, a database query, or a temporal reference. Its premise is to support standard, unambiguous linking and rendition of any kind of document or format.

3.1 A Brief History of Hypertext

Using a criteria defined by Conklin in his survey [Conklin86], the concept of hypertext can be characterized as linking supported by a tool. Vannevar Bush's imaginary machine memex (for memory extender) can thus be considered a precursor to the computer-based hypertext systems that came about in the 1960s. Bush contentions was that Man thinks primarily by association, and that support for this kind of intellectual work could be realized in the form of a machine. The memex that he envisioned bears striking resemblance to the computers on our desktop today. He also foresaw an information age, where "trail blazers" would offer their services to provide paths in the vast oceans of available information [Bush45].

The Bush article, published in 1945 (authored in the early 1930s, Bush was reluctant to publish it) was to be influential on the design of hypertext systems realized two decades later. Douglas Engelbart and his team at the Stanford Research Institute created NLS (the oNLine System) and Augment, where all information such as e-mail, reports, files etc. were stored and accessible electronically, while Ted Nelson (who coined the term "hypertext") published his vision of a global, interconnected information space.

Engelbart has justly been called the Edison of our times, for having invented the mouse, computer supported cooperative work, and many other tools and techniques that have since become commonplace. Engelbart's team made a fundamental contribution which other early hypertext designers missed entirely, which was the separation of links from their endpoints or anchors. By maintaining this separation, it became possible to reference documents through addressing: all you need is a location. In Engelbart's system, files were broken down into statements (small chunks) being addressed by a hierarchical, alphanumerical addressing convention, whereas subsequent hypertext systems were modelled using only the concept of nodes and links. The granularity of links endpoints became thus dependent on the choice of node for these systems.

The hypertext systems that were to evolve for the next two decades all failed commercially. With hindsight, it is clear that the problems that plagued them were due to a combination of factors. Firstly, the system were monolithical: all of the documents needed to be controlled and accessed through a proprietary application. Secondly, one could not export or import data while preserving the link information. Thirdly, because of the granularity, the design of the systems were not optimal for every kind of use—some systems would lend themselves to page-sized chunks of information (e.g. KMS), other systems (such as NoteCards) provided a document model of cards within cardboxes. These early attempts pinpointed two problems inherent with the medium:

- Authoring hypertext is hard, because of the cognitive overhead of authoring: premature structuring requirements and the necessity of addressing multiple reading paths.

- Reading hypertext is hard. You can quickly become lost in hyperspace.

The IRIS team at Brown University reintroduced the concept of anchors in the mid-80s, in addition to nodes and links, and pioneered the separation of link data as an externally stored web [IRIS90]. They also suggested adding links as a standard part of operating systems, to help applications to express links in a unified, portable manner.

The Dexter Hypertext Reference Model was one of the first attempts to formalize a model of hypertext. At about the same time, it was realized that SGML is an effective means of representing hypertext documents portably [Coombs87].

3.2 Hypertext as Structured Documents

HTML, the markup language of the WWW, strongly resembles SGML because of the use of a tag naming scheme and similar markup delimiters. HTML can be viewed as an application of SGML, where the tags are fixed, and the linking semantics reside outside of the scope of validation (as the named links cannot be validated by SGML, though they can be executed and understood by SGML-based applications). Originally created without a DTD in mind, HTML is migrating towards conformance with SGML. The WWW has however yet to become a fully-fledged SGML application, as the great majority of documents on the WWW have been created without SGML validation.

Perhaps the greatest virtue of HTML has been to expose generic markup techniques for a global audience, while focusing on simplicity. Simplicity is hard to achieve in a pure SGML-based system: SGML addresses platform independence through provisions for character sets, removing file system concepts through entities, and providing other insulating and general abstract solutions—complexities that HTML has simply sidestepped. In addition, SGML is marred by optional features designed to simplify data entry, which have become unnecessary with the advances in user interfaces.

One may be lured into thinking that HTML and other de facto standards may suffice. SGML and its related standards are however necessary in adequately addressing Integrated Open Hypermedia.

4. Integrated Open Hypermedia

The SGML formalism uses freely named tags to encapsulate data as elements. As tags are non-disjunct or contain one another, the markup creates a hierarchy of elements, which also reflects the granularity of the encoded information. SGML's document type definition, or DTD, describes a class of documents, for which various permissible instantiations can be made. As SGML documents are validated against their DTD, the SGML approach guarantees a neutral conformance with a methodology that provides for hierarchy and element granularity. These qualities are key points in understanding why SGML, and SGML-based approaches, so well address the field of hypermedia.

As the DTD is designed for classes of documents, the knowledge representation of documents in SGML is met by a domain-specific, and extensible, set of tags. As one assembles larger and larger document collections, it is imperative that the document contents remain accessible for queries. Content-oriented tagging—where the tag naming scheme addresses primarily usage and meta-information and not layout concerns—is a sound approach to document storage.

As SGML by nature is hierarchical, it is well suited for lenghty and technical documents. In contrast, HTML has a limited tag set and a flat document markup—leading to a poorer encoding, and chunking of information, as existing HTML browsers do not have adequate navigation features for large documents. As an example, a set of Norwegian government whitepapers are published on the WWW—the SGML source does not have a single link; the generated HTML version has in excess of 10,000 links, to compensate for the lack of hierarchy and corresponding navigational support in HTML browsers.

Unfortunately, the document model of SGML is limited to the realm of a single document; though SGML is perfectly adequate to represent e.g. links to other documents, or complex multi-media documents, software based on the SGML standard alone cannot verify or implement the semantics behind the linking. SGML-based approaches such as TEI extended pointers, and HyTime—the international standard for hypermedia—extend the use of the SGML standard to cover these requirements.

4.1 HyTime

The intellectual heritage from the field of hypertext pinpoints the need and usefulness of separating links from anchors, and of being able to store this information within documents, and externally as well. Also, documents need to remain application and platform independent. Incidentally, though HyTime meets all of these objectives, it does so from being originally a standard proposal for the description of music [SMDL]—which has since become an application of HyTime.

HyTime is a modular standard, where the highest module deals with the rendition of time-based documents. It is possible to use (and implement) only portions of the standard. In contrast to SGML, which mainly addresses issues of structure without dealing with semantics, HyTime provides a vast number of pre-defined templates for hypermedia.

The requirements of specific tag names has been avoided through the use of architectural forms (enabling architectures) that permit HyTime engines to recognize the templates regardless of the tag naming scheme. HyTime is also extensible; in object-oriented parlance, it provides for multiple class inheritance and extensibility.

4.1.1 HyTime linking

HyTime defines but two link constructs, one of which is a special case of the other. The richness of HyTime linking lies in the many ways one can describe locations in document, often as a sequence of stepwise refined addresses known as location ladders.

The contextual link or clink (pronounced c-link) is the simpler of the link constructs. One of its anchors ("the context") is part of the document where the link markup resides. The other endpoint can be in the same or somer other document. The clink is rather common as it aptly describes normal cross-references.

The independent link or ilink (pronounced eye-link) permits the link data to be stored externally, separate from the document(s) that the link markup connects. One can thus make changes to the link data without modifying the documents being linked.

4.1.2 HyTime location addressing

Locations can be defined using different addressing methods in HyTime. Three basic location addresses are:

- Name Location (by an SGML name)
- Tree Location (by SGML tree structure, thanks to the markup encapsulation)
- Data Location (by any type of data, counting e.g. words)

In contrast, HTML can at its best mimic the usefulness of name location addressing.

4.1.3 Dynamic Linking and Assembly

Information today is rapidly becoming a transitional asset, where documents may be assembled on the fly, say, as the result of a query. HyTime also covers the dynamic handling of such documents: link resolution is performed when the HyTime application requests it—typically as the result of an end-user selection. This late binding permits the gradual creation of interlinked document collections, where some documents may exist only as a query specification. Dynamic addressing can also be done through queries, where the exact location is not known but resolved through matching against some element properties, in real time. HyTime defines a query language (HyQ) which is now, through ISO ruling, superseded by the DSSSL (Document Style Semantics and Specification Language) standard document query language SDQL. Thus SGML, HyTime, and DSSSL have all been brought into alignment for consistency during 1995.

An interesting addition to the WWW is the emergence of indexing techniques that tap into this docuverse. As a side-note, it is worth noting that there are far more documents in SGML than currently in HTML. In fact, there is (at least) one single SGML document that contains more textual data than are now in HTML on the WWW.

4.1.4 A few typical uses for HyTime

The use of the addressing mechanism outlined above, with clinks, permits e.g. to attach annotations to documents to which one only has read-only access. Similarly, one could use the same features for electronic document review. Because of the strength of stepwise addressing, using location ladders, addressed documents may even be edited without the links being necessarily affected adversely.

Such uses of HyTime are already supported by commercial applications, including an SGML-based browser for the WWW.

As HyTime becomes implemented to a larger degree—which will take a while—it may take some time before we see all of its possibilities becoming available. HyTime has already attracted much attention for the design of interactive training manuals and for IETMs (Interactive Electronic Training Manuals), a problem domain which is funded rather forcefully.

5. Addressing Publishing and Layout

Being a neutral standard, SGML lends itself well to publishing on multiple media from SGML source. The conversion process is typically a filter that generates formatter output as part of the tree traversal of the SGML document. The formatted output of an SGML document has heretofore been addressed mainly through application specific solutions; the recently adopted Document Style Semantics and Specification Language is likely to become rapidly implemented, at least for specifying on-line viewing. However, as formatting specifications often are contradictory, uses for which page fidelity are necessary may best be handled by proprietary solutions such as PDF. Along the same vein, documents with particular layout concerns—such as poetry and math—are currently likely to be published more efficiently using traditional techniques.

5.1 The Strindberg project

As a final example we will mention an ongoing work with SGML-encoding of the Swedish author Strindberg's writings. As part of evaluating the TEI guidelines, we have at the Royal Institute of Technology (KTH) marked up two most recent volumes of the critical National edition of August Strindberg's collected works according to the TEI guidelines. An initial markup was done from ASCII WP files using Author/Editor and MS Word. These files were then edited and validated using emacs and OmniMark, respectively. An OmniMark script produced input for the LaTeX formatter. Finally we delivered PostScript files to the printing house. Both volumes—*Svarta fanor* and *Dikter på vers och prosa*—were published in late 1995 by Norstedts, Stockholm.

Up to quite late in the production, we were able to produce hard copy directly from the SGML source. However, as we approached publishing deadlines and needed to address delicate layout concerns (widow control, insertion of graphics, etc), we eventually made the transition to edit the LaTeX file directly. Final hard copy was in the form of PostScript files to the typesetter.

A major problem—for which we underestimated the necessary resources for—was the minutious work necessary to mimic and duplicate the typographic conventions used in this series (there were 35 previously published volumes). One of the volumes was poetry, which had its own peculiarities regarding line breaking, hanging indentation, optical centering (in contrast to true centering) and other layout-related issues that greatly complicated the whole process. In addition, we needed to acquire special fonts (such as ancient greek—plenty of accents), make changes to the placement of diacriticals in the fonts to match earlier volumes, edit the kerning table, and create a PostScript file to render a rune glyph inline.

As the primary goal was to address publishing, the granularity of the markup was about as coarse as needed to adequately differentiate elements typographically. More work could obviously have been done to capture nuances in Strindberg's work to apply more of the TEI ideas. We are currently planning such endeavours. As a prelude we have produced a very rough and simple TEI-encoding of all the 37 volumes now in print.

6. Conclusion

We have yet to achieve Vannevar Bush' memex. Similarly, Engelbart's augmentation of the human intellect through computer supported cooperative work is very much in its infancy. The emergence of a family of international standards for Integrated Open Hypermedia however hold the promise of these visions as they are, and remain, enabling technologies.

Glossary

Hyperlink: An information structure that represents a relationship among two or more objects. [HyTime, ISO/IEC 10744:1992, clause 3.15]

Hypermedia: in computer science, the integration of graphics, sound, video, or any combination into a primarily associative system of information storage and retrieval. Hypermedia, especially in an interactive format where choices are controlled by the user, is structured around the idea of offering a working and learning environment that parallels human thinking—that is, an environment that allows the user to make associations between topics rather than move sequentially from one to the next, as in an alphabetic list. Hypermedia topics are thus linked in a manner that allows the user to jump from subject to related subject in searching for information. For example, a hypermedia presentation on navigation might include links to such topics as astronomy, bird migration, geography, satellites, and radar. If the information is primarily in text form, the product is hypertext; if video, music, animation, or other elements are included, the product is hypermedia. [Microsoft (R) Encarta. Copyright (c) 1993 Microsoft Corporation. Copyright (c) 1993 Funk & Wagnall's Corporation]

Multimedia: the combination of sound, graphics, animation, and video. In the world of computers, multimedia is a subset of hypermedia, which combines the elements of multimedia with hypertext, which links the information. [Microsoft (R) Encarta. Copyright (c) 1993 Microsoft Corporation. Copyright (c) 1993 Funk & Wagnall's Corporation]

References

Broady, Donald (1995): “Det nya handbiblioteket”. In: *Biblioteken, Kulturen och den sociala intelligensen* (Ed. Lars Höglund). Göteborg: Forskningsrådsnämnden/Valfrid, 1995, 83-107.

Bush, Vannevar (1945): “As We May Think”. *Atlantic Monthly*, July 1945, 101-108.

Conklin, Jeff (1987). “Hypertext: a survey and introduction”. *IEEE Computer*, 20 September, 17-41.

Coombs, James H., Renear, Allen H., and DeRose, Steven J. (1987) “Markup Systems and the Future of Scholarly Text Processing”. CACM, 30 (11), 933-947.

Engelbart, Douglas C. and English, William K. (1968). “A Research Center for Augmenting Human Intellect”. In Proceedings of the Fall Joint Computer Conference.

Goldfarb, Charles (1990). *The SGML Handbook*. Clarendon Press, Oxford.

Haitto, Hasse (1995). “SGML, HyTime, and the WWW”. Proceedings of SGML Europe ’95, Gmunden.

HyTime (1992). Hypermedia/Time-based Structuring Language ISO/IEC 10744:1992, International Organization for Standardization.

[IRIS] Institute for Research in Information and Scholarship, Brown University (1990): *User’s Guide: IRIS Intermedia, Release 3.0*, Providence, RI.

Nelson, Theodor H. (1987) : *Literary Machines*, Edition 87.1.

[DSSSL] ISO/IEC 10179:1996 Information Technology — Text and Office Systems — Document Style Semantics and Specification Language (DSSSL), 1996.

[SGML] ISO/IEC 8879:1986 Information Technology — Text and Office Systems — Standard Generalized Markup Language (SGML). International Organization for Standardization, 1986.

[SMDL] ISO/IEC CD 10743 Information Technology — Standard Music Description Language.
International Organization for Standardization.

[TEI] Burnard, Lou/Sperberg-McQueen, C.M (Eds.): *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*. Volume I-II. Text Encoding Initiative, Chicago University/Oxford University, 1994.