Just In Time Markup for Electronic Editions

Phillip William Berrie
Australian Scholarly Editions Centre
Australian Defence Force Academy
AUSTRALIA
p.berrie@adfa.edu.au

Abstract

This paper describes a new paradigm for the creation and maintenance of electronic editions of works of historical literature that use the Standard Generalised Markup Language (SGML) to encode their markup. The Just In Time Markup (JITM) paradigm solves a number of common problems identified in the systems currently in use for creating electronic editions. The paper also shows how the nature of the JITM paradigm will reduce the significant workload involved in creating and authenticating an edition while allowing the edition to be a continually evolving research work for any number of scholars — even after publication.

Introduction

In the field of the humanities, information technology has found application in the creation of electronic transcriptions of historical documents. The original documents may be both rare and fragile and therefore not readily accessible to scholars who wish to study them[i]. Authenticated, electronic editions not only give scholars cheap and easy access to the content of these documents, but also allow them to use computer-based tools for studying the material's textual content.

The standard markup language syntax for these editions is that defined by the Document Type Definition (DTD) of the Text Encoding Initiative (TEI) [MLA, 1997]. The TEI DTD [see Sperberg-McQueen et al., 1994] is an application of SGML[ii]. In the preface of Charles Goldfarb's book, *The SGML Handbook* [p. xii Goldfarb, 1990], he likens people's understanding of SGML to that of blind men examining certain parts of an elephant and coming to a limited understanding of the nature of the beast. To stretch the metaphor further, the development of the TEI DTD has made the elephant manifest in all its glory so that even sighted men must be in awe, and daunted at the task of making the beast do some work — and pity the poor blind men.

It is a contention of this paper that these complexities have slowed down the use of electronic editions in the humanities. The only institutions adopting these new technologies have been scholarly institutions, which have developed their cross-discipline skills enough to be able to tag literary meta-data in SGML in accordance with the TEI DTD. This is not necessarily a bad thing, but the author's experience indicates that the overheads involved are not inconsiderable and that the combined skills of literature scholar and SGML encoder are not readily found in the same person. Simpler methodologies and tools are required to reduce the overheads involved in developing and maintaining electronic editions.

Why are Electronic Editions Different?

An electronic edition can be defined as "an enhanced transcription of a document in the digital medium". Three aspects of this definition make the creation and maintenance of electronic editions different and more difficult than that of normal electronic documents.

The first and most important aspect of the creation of an electronic edition is that it is a transcription of an existing document. The accuracy of the transcription process in the creation of any new edition, either physical or digital provides the basis for the edition's authority as a scholarly resource[iii]. This requirement for accuracy increases the preparation time for a scholarly edition greatly[iv].

SGML-based electronic editions compound this problem because their markup is embedded into the text of the transcription. Therefore any change to the markup potentially can change the transcription and therefore needs to be checked thoroughly.

The second aspect involves the addition of enhancements into the electronic edition. The implementation of these enhancements involves the inclusion of such features as structural elements for navigation; ancillary material to provide extra context for understanding the work; and editorial comment to provide authority for editorial decisions.

Limitations in the nature of SGML mean that important decisions about structure and meta-data content have to be made during development of the edition to make the edition useful to as many people as possible. These decisions can bias the use of the edition against people with requirements that were not catered for in the original design.

Finally the ease with which digital documents can be invisibly changed and easily propagated over computer networks makes the continued reliance on the authority of an electronic edition questionable. The post-publication authentication of electronic editions is not very strong in currently available editions. Most of them rely on the edition's files being kept on non-volatile media such as CD-ROM for their protection. However since the advance of technology may make both the current CD format and its reader technology obsolete this is not a good long-term solution[v]. Document authentication technologies which were recently recommended in the *Guidelines for Electronic Scholarly Editions* [MLA, 1997] can be used to guarantee the authenticity of the edition no matter where it is stored, but use of the suggested invasive technologies (eg. digital signatures) completely lock electronic editions into the same paradigm as their physical predecessors - the printed book.

The Just In Time Markup Paradigm

The ideas behind the JITM paradigm came out of the author's association with the Australian Scholarly Editions Centre (ASEC) at the Australian Defence Force Academy[vi]. Having assisted with the preparation of a number of physical scholarly editions the author was very much aware of the time consuming nature of such a task. With the decision of the Centre to experiment with electronic editions as a means of increasing the availability of scholarly resources in the area of historical Australian literature the author was lead into an investigation of the current tools available and found them wanting.

It is the author's view that the most important part of creating an electronic edition is the accurate transcribing of the original work into the new medium. All other data associated with the electronic edition is considered subjective or optional and is abstracted away from the transcriptions to be applied as required.

What is Just In Time Markup?

A JITM system allows a user to create on demand user-customised versions of electronic editions of historical documents. In a JITM system the specifications for SGML markup of meta-data for the document is stored separately from the rigorously authenticated transcription of the document and specific markup is inserted into the transcription file by the JITM system prior to being viewed by an SGML-capable browser.

Why Just In Time Markup?

By abstracting the meta-data markup away from the transcription files of the edition we ensure that the transcriptions remain authentic removing a lot of the proofreading burden of creating an edition. Secondly, creation of the meta-data for the edition does not need to be done concurrently with the creation of the transcriptions. Meta-data can be created and applied to the edition incrementally, concurrently and even after publication of the transcription files. The paradigm will also support sets of conflicting meta-data (although not applied simultaneously). This means that the electronic edition can become an evolving work of scholarship based on the work of many hands, which should negate the need for the creation of any future editions in the electronic medium.

How does it work?

The JITM paradigm defines the transcription of the historical work to be a text defined as follows;

A text is a sequence of words, punctuation (and white space) represented by characters and entity references.

The term 'entity reference' in the definition is an SGML mechanism for representing characters not supported by the ISO 646 character set[vii]. This definition excludes reference to the typographical features of the original source such as page layout and emphasis[viii]. This is done purposefully. In the JITM paradigm both these aspects of the work are abstracted away from the transcription as they could bias the use of the transcription file.

A JITM-based system relies on the fact that once the transcription of the original has been performed, and the resulting file authenticated, the textual content of the file should not change. The static nature of the transcription file allows us to define a reference scheme (based loosely on the HyTime Standard [ISO, 1997]) for the insertion of markup into the transcription file when required to create a "Perspective" of the edition. In the JITM paradigm a perspective is a temporary document which combines the transcriptions of the original work with a selected set of meta-data specified by the user.

The insertion of meta-data markup is done by iteratively converting appropriate sections of the transcription file into linked lists at the time of instantiation of the perspective.

The elements of the list are strings of characters from the transcription at various granularities (i.e.. paragraphs, words and characters). The following diagram demonstrates the procedure graphically at the word token level.

transcription file word token stream

transcription file word <i>token stream</i>
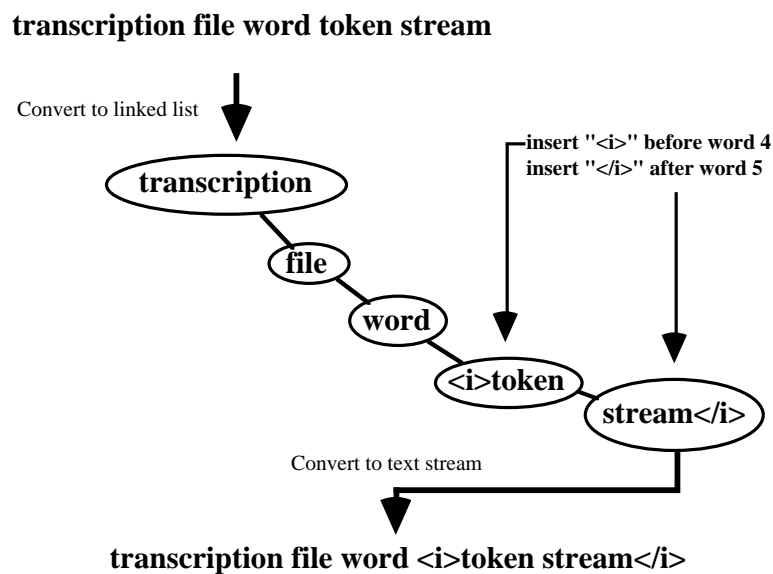
Figure 1. Tokenising of text string

If necessary specific elements of a list are broken down further for insertion of markup at a finer level. This can cause the creation of lists up to three levels deep where a tag is to be inserted into the body of a word. In cases where tags should be embedded at the same position the order of insertion is determined by stack data structure which ensures the latest opened tag is closed first.

The use of a linked list means that modification of a list element does not change the position in the list of other elements as would happen with inserting the markup into a string of characters. As the underlying data format for the transcription file is basically text, this implies that all markup needs to be added in a single pass, as the linked list for a modified file may be different. This requirement is in fact beneficial to the authentication process, as it ensures that a set of meta-data insertion commands that would work for a specific transcription file would most likely not make sense when applied to a modified version of the file. It is a simple process to convert the linked lists back to a contiguous stream of text for output.

Content Authentication

As mentioned earlier, the proposed JITM paradigm depends on the transcription files being static so that the meta-data markup can be inserted into the correct places in the transcription file. To make sure this occurs, a JITM system needs to be able to check that the transcription file to which the meta-data is to be applied is identical to the one used for its creation. If this is not the case the markup may be inserted incorrectly. This means that the authentication of the transcription files is essential to the correct operation of a JITM system.

The JITM paradigm's authentication scheme continues the abstraction theme used for the rest of the paradigm. The transcription files are not altered in anyway to support the authentication scheme. Therefore the files cannot possibly be altered by this scheme. With invasive authentication schemes, such as digital signatures, new versions of the files would have to be created if there were a change in software, algorithms, or even document ownership, thereby creating new states of the transcription files with all the disadvantages that entails for future scholars[ix].

The authentication mechanism used is very simple compared to the more advanced document authentication schemes mentioned above. However, since it is only used to guarantee that the transcription files have not been modified, it should be sufficient for the task. The scheme uses the early cryptographic concept of a Manipulation Detection Code (MDC) [p. 9, Christoffersson et al., 1988]. In practice the MDCs are calculated for suitably sized text elements of the transcription files (eg. paragraphs for prose) using a publicly known one-way hash function[x]. The MDCs for a text element are incorporated into the detail of meta-data that can be applied to that text element. It is an illegal action for a JITM system to embed tags into a text element where the recorded MDC does not match the MDC calculated from the text element. Therefore the user is immediately made aware that some element of the JITM system in use is not authentic.

This leads us to a subtle strength of the JITM paradigm. One of the paradigm's advantages is that users can develop their own meta-data and swap them to increase the range of perspectives that can be generated for a document. Since the MDC stored with the meta-data is that of the transcription file used to create it, imported meta-data can be used as a quality check for other JITM systems. If the MDC stored in a piece of meta-data proves to be invalid when used on another copy of the transcription file, but is valid on the system that created it, then one of the systems has been corrupted. The users can then discover which files are wrong and replace them from an archival source. This process should prevent the proliferation of corrupt states of the transcription files within the community.

Implementation of the System

This section looks at how a JITM system could be implemented so that it would complement existing SGML-based tools for electronic editions. To promote acceptance of the JITM paradigm it should impact the user's environment as little as possible while providing a greater versatility and proof of authenticity than existing systems.

The most important aspect of the JITM paradigm is that the perspective documents that the system delivers are virtual documents created on demand. This means that prior to viewing a JITM perspective the user must give the system some parameters that will be used to create the perspective.
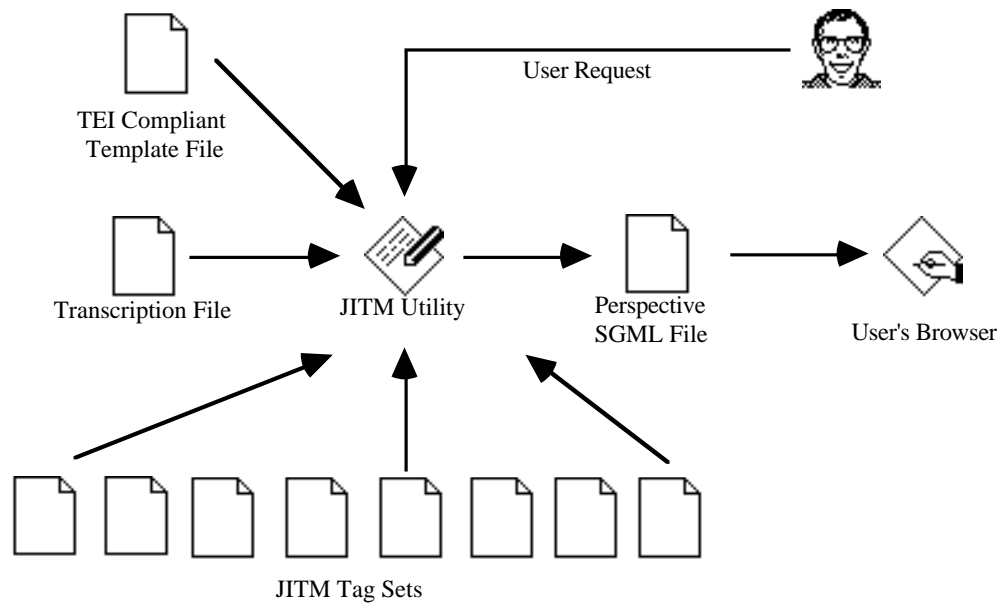
Figure 2. Implementation of JITM System.

As shown in the diagram above this could be achieved by having the user access a front-end application (the JITM utility). This application would allow them to choose the transcription file and the aspects of the work in which they are interested. The JITM utility would then generate the appropriate SGML file, which would be delivered to the user's SGML-browser. In fact any number of perspective views could be viewed simultaneously on a browser capable of multiple windows allowing the user to compare the different ways of looking at the work under study.

This design allows the system to be both stateless and scalable. Once the JITM utility has created the perspective document it does not need to remember anything about the user's request. The system is therefore suitable for implementation either as a stand-alone application on a personal computer or run as a Common Gateway Interface (CGI) application on a web server. The prototype that is demonstrated in this paper uses a web-based delivery system for users of the edition. Stand-alone systems would be part of a suite of tools required for creators of meta-data for the edition.

Benefits of the JITM Paradigm

One of the biggest strengths of the JITM paradigm is its modularity. This gives benefits at all levels of the production of an electronic edition both pre- and post-publication. The following are some of the advantages of the JITM paradigm:

- Transcription files only need to be authenticated once, the JITM system then continually checks that the transcription is still authentic no matter how much meta-data is added to the edition.
- Different logical structures can be applied to the transcription files to suit different tasks. For example, markup for the typography of the original document can be applied if required (i.e. lineation and pagination) and ignored if not wanted.
- Conflicting logical markup can be developed for an edition and applied selectively depending on the user's requirements. Browsers that allow multiple windows could be used to allow these conflicting structures to be visually compared.
- Editors can develop meta-data in isolation from other potentially distracting aspects of the work reducing the complexity of this task.

- The end user can apply the work of many different editors to create perspectives specifically designed for their needs.
- If necessary users can generate their own meta-data after publication of the transcription files.
- The authentication mechanism of the JITM paradigm prevents the propagation of variant versions of the transcription files by providing an authenticity check when meta-data is transferred between sites.

Finally the transcription files consist only of the transcribed text of the original document with minimal markup for identification purposes. This means that it is very easy for these transcription files to be re-use for other purposes.

Example of JITM in Action

The potential capabilities of a JITM system seem very great. However to the uninitiated it is sometimes hard to come to grips with what the JITM paradigm allows. To help explain the system here is a simple example of how different perspectives can be generated using a JITM system.

The following text is a single text element from a transcription file.

<P id="p-4">

"It's only a nominal thing, old man," Frere said to his old comrade, when they met. "That parson has made meddling, and they want to close his mouth."

</P>

The following are two sets of simplified markup instructions (i.e. tag sets) for this paragraph.

<tagRecord tid="p-4" key="652" pos="1.001">LB</tagRecord>

<tagRecord tid="p-4" key="652" pos="13.001">LB</tagRecord>

<tagRecord tid="p-4" key="652" pos="22.001">LB</tagRecord>[xi]

<tagRecord tid="p-4" key="652" pos="1.001">speech speaker="Frere"</tagRecord>

<tagRecord tid="p-4" key="652" pos="7.004">/speech</tagRecord>

<tagRecord tid="p-4" key="652" pos="17.001">speech speaker="Frere"</tagRecord>

<tagRecord tid="p-4" key="652" pos="28.006">/speech</tagRecord>

The original creator of the edition would have generated the first set, as they would have transcribed the text from an original source document. At this point it would be a simple task to record such typographic information[xii].

The second set could have been created at any time after publication of the authenticated transcription files. In this case it may be a linguist who is interested in analysing the language usage of the characters in the work. This may be something the original creators of the edition did not consider to be of interest.

Calculating the MDC for the text element yields a result of 652 therefore it is valid for a JITM system to apply the specified tag sets to this text element as the MDCs match. Applying the first set would yield the following result

<P id="p-4">

<LB>"It's only a nominal thing, old man," Frere said to his old <LB>comrade, when they met. "That parson has made meddling, <LB>and they want to close his mouth."

</P>

Applying the second set to the original text element would yield the following result.

<P id="p-4">

<speech speaker="Frere">"It's only a nominal thing, old man,"</speech> Frere said to his old comrade, when they met. <speech speaker="Frere">"That parson has made meddling, and they want to close his mouth."</speech>

</P>

Note: Applying the second tag set to the text element modified by the first tag set would not be legal as the MDC for the modified text element would not be the same as that stored in the records of the tag set.

Applying both tag sets together in the same pass is legal and would yield the following result.

<P id="p-4">

<LB><speech speaker="Frere">"It's only a nominal thing, old man,"</speech> Frere said to his old <LB>comrade, when they met. <speech speaker="Frere">"That parson has made meddling, <LB>and they want to close his mouth."</speech>

</P>

This modified text element might appear on the screen of a browser something like the following.

"It's only a nominal thing, old man," Frere said to his old comrade, when they met. **"That parson has made meddling, and they want to close his mouth."**

Here the bold character style has been used to differentiate the dialogue in the text element based on the underlying markup to demonstrate the differentiation provided by the markup. Although a simple demonstration this use of tagging could be used to highlight the dialogue in a perspective that uses the lineation of the original source document.

To enumerate the potential number of usages of an electronic edition created using a JITM system the number of different perspectives available for an edition is limited by the following equation.

Number of Possible Perspectives = $2^{\text{(No. of tag sets)}}$

Note: Limitations of the paper-based format of this paper have forced the simplification of the tag set information. The information left out includes information about how the text should be inserted and information used to help correctly parse the document against the TEI DTD in the case of coincident tags.

Problems with the Paradigm Shift

This section looks at problems foreseen with the acceptance of the JITM paradigm. These problems are concerned with the modularisation of the electronic edition under the JITM paradigm and have to do with intellectual property.

Issues of Ownership

The JITM paradigm logically separates the original transcriber's editorial content from the transcription. The ownership of the component parts of an edition now becomes an issue. There is a strong case that the transcription files, being an exact transcription of the original state, may not be considered sufficiently altered by the transcriber[xiii], for them to be able to claim their own copyright on the material. In this case, the transcription files would either be in the public domain or owned by the holder of the copyright of the original author and therefore the creator of the edition would not be able to claim royalties for the transcription files. Traditional methods for creating editions add extra material into the text of the transcription such that the editor can claim their own copyright on the finished product.

It follows that the creator and or publisher of the edition may only be able to claim copyright over their meta-data, and therefore may only be able to charge royalties for the use of such. Since the selection of meta-data is optional in a JITM system, this means that the meta-data of the transcriber could be replaced with somebody else's work and the creator would then have no claim for recompense over the users of the transcription file.

The perspective documents created in a JITM system are another area of concern with regard to copyright issues. Since a perspective is a transient document created for displaying particular aspects of the edition, and is temporary in nature, could this fall under fair usage for scholarly research and therefore be exempt from any claim for royalties? What happens in the case where a perspective is generated from meta-data from a number of different sources? How is the copyright of a perspective attributed when it is an amalgam of different people's intellectual content?

Demoting the Editor

The JITM paradigm's devolving of the responsibility for the edition away from a specific editor could either "make or break" the acceptance of the new paradigm by the scholarly editors themselves. After all, the paradigm makes the transcription files the most important element of the edition, reducing the editorial contribution to a set of optional files that can be replaced if desired. While this facility of the JITM paradigm may be good for the edition, in that it allows the edition to be a continuously developing resource, it does down play the contribution of the original creator of the edition.

This demotion of the original creator to the level of just another supplier of optional meta-data could find resistance amongst scholarly editors who perceive the creation of an edition as a way to immortalise themselves "in print". Also, considering the amount of work that goes into the creation of a fully annotated critical edition, it is understandable that any editor could become discouraged to realise that all their efforts can be ignored by the user of the edition.

These are all reasons why the more traditional methods for creating electronic editions might be favoured over a JITM system by potential edition creators. However the reduced time to "publication" of the transcription files, and the unlimited extensibility of a JITM edition, will hopefully make the paradigm attractive to those editors who feel limited by the current systems, which are still bound by the same limitations as a physical edition.

Further Work

This section looks at the various avenues of investigation that are being considered for the further development of Just In Time Markup.

Management & Manipulation of Meta-data

A JITM system has two major types of persistent data, the transcriptions and the sets of meta-data. The content of both data types is defined so that they are very portable. This enables the easy repurposing of the files of a JITM system. Transcription files are specifically defined so that they include little extraneous text so that they can be used in other systems. Similarly, meta-data stored as files can have their own uses. Users of a JITM system can transfer meta-data without having to transfer transcription files because the JITM paradigm ensures that the application of the meta-data authenticates the transcription files.

The highly structured nature of the meta-data file format makes them eminently suitable for storage in a database. This would allow the meta-data to be stored in a more efficient manner and help reduce redundant data so that global changes could be made efficiently. The current JITM design associates meta-data directly with the section of the transcription it refers to, but helps little with handling global interactions. Maintaining the meta-data markup in a database would enable such global interactions.

This leads to another intriguing aspect of maintaining an edition's meta-data separate from the transcription file. For some purposes the meta-data itself is all that is required. The prime example is analysing whether a chosen set of meta-data will create a valid SGML document conforming to the TEI DTD. The JITM paradigm allows the development of meta-data that may use conflicting structural elements, but the TEI DTD restricts a perspective to a single hierarchy.

The meta-data could be parsed in isolation against the DTD before it is embedded into the transcription file. Another example of the use of separate meta-data is based on the area of statistical analysis of a particular type of meta-data where it is more the existence and attributes of the meta-data that is important rather than the textual content.

Despite the advantages of using databases for storing JITM meta-data it is important that any JITM system maintains the capability to output its data types in the simplest of data formats. By using the lowest common denominator approach this will in turn prolong the usefulness of the information for archival purposes. Tying the meta-data to some proprietary data format would possibly limit its long-term usefulness.

Delivery Mechanisms

The usefulness of SGML-based electronic editions has been somewhat constrained by the lack of cheap SGML browsers[xiv.] Access to a lot of electronic editions has been made available through the World Wide Web by converting SGML to HTML for presentation on the freely available web browsers. However the limited capabilities of HTML make this unsatisfactory for anything other than reading.

The eXtensible Markup Language (XML) is being touted as the replacement to HTML, and is also supposed to bring some of the power of SGML to the World Wide Web. A long-term JITM-based edition would need to be able to support the use of an XML-based delivery system to ensure its continued use. Fortunately there are a few things in favour of the two systems being compatible, and potentially it is possible that an XML browser could replace the SGML browser currently required for a JITM system. Work is planned to look into using XML as the delivery mechanism for a web-based JITM system.

Conclusions

The process of creating an electronic edition of a historical document is most importantly an act of transcription of the document into the digital medium. If the electronic edition is to have any scholarly authority the users of the edition must have a certain level of trust that the original transcription was accurate and that the authenticity of the transcription in the edition has not been compromised since it was created. An electronic edition based on the Just In Time Markup paradigm depends on the unchanging nature of the transcription files for its operation and, therefore has a built-in authentication mechanism, which continually verifies that the files to be marked up are exact copies of the original transcription.

The definition of the JITM paradigm fares well when compared to requirements for electronic editions as listed in the recently released *Guidelines for Electronic Scholarly Editions* [MLA, 1997]. However despite its advantages it would appear that the main problems for a JITM system would be one of acceptance. The JITM system is very different to the current paradigm, which is based more on the precepts of the printing of physical editions. A JITM-based edition is not likely to be a money making venture and so its maintenance and distribution would necessarily become the full responsibility of the editor up until the point where the JITM paradigm is accepted by the scholarly community. Fortunately the digital nature of the edition makes this easy using World Wide Web technology.

Furthermore, unlike the currently available editions, which become frozen at the time of publication, an electronic edition created using the JITM paradigm can continue to evolve as more meta-data is developed. This ability to continually evolve plus its non-reliance on proprietary standards will hopefully future proof a JITM-based edition so that it will remain useable for a long time. Print-based editions can last hundreds of years, it remains to be seen whether electronic editions can last as long. It is the belief of the author that the best way to "keep" an electronic edition is through use. By making an edition as reusable as possible using the JITM paradigm users will propagate the edition onto whatever computer systems they will use in the future. Hopefully this will make the maintenance of the edition a collaborative activity and keep the edition useful beyond the life span of any one individual or institution.

Bibliography

Christoffersson et al., Crypto User's Handbook, Elsevier Science Publishers B. V., 1988.

Goldfarb, C. F., *The SGML Handbook*, Clarendon Press: Oxford, 1990.

International Standards Organisation, *Hypermedia/Time-based Structuring Language (HyTime)*, URL: "http://www.ornl.gov/sgml/wg8/docs/n1920/html/n1920.html", 1997

Guidelines for Electronic Scholarly Editions, Committee on Scholarly Editions of the Modern Language Association of America, URL: "http://sunsite.berkeley.edu/MLA/intor.html", 1997.

Sperberg-McQueen C. M. and Burnard L. (Ed.), *Guidelines for Electronic Text Encoding and Interchange,* Text Encoding Initiative, 1994.

This paper is based on sub-thesis written by the author to fulfil the requirements of a Masters of Information Science through the School of Information Science at the Australian Defence Force Academy. Further information on the JITM paradigm can be obtained from the Australian Scholarly Editions Centre's web site at "http://www.adfa.edu.au/ASEC/".

Endnotes

i   The Dead Sea Scrolls are a perfect example of this. The originals are not only unique, but are so badly deteriorated that scholars are only allowed access to facsimiles of the original scrolls.

ii SGML is a metalanguage for defining markup languages that can be used to describe the logical structure of documents. Markup languages can also be used to add meta-data without interfering with the readability of the text through the use of hypertext links.

iii One of the major criticisms of Project Gutenberg by scholarly editors is that the accuracy and provenance of the transcribed texts is insufficient for proper scholarly research, but its ready availability means that it could be used as such.

iv In a survey done by the author [see authors sub-thesis] it was discovered that about 50% of the preparation time for a scholarly edition is concerned with accuracy of the transcription.

v   These problems are in addition to the unknown durability of the CD storage medium.

vi The ASEC has produced a number of physical scholarly critical editions of famous 19th Century Australian literature.

vii The ISO 646 character set is the default character set for SGML applications and is a subset of the standard ASCII character set. It is left up to the user's browser to display the characters represented by entity references.

viii Lack of specific character representations for emphasis is a limitation of computer encoding systems. Even the two-byte Unicode scheme does not have separate characters for bold or italic versions of characters.

ix Digital signature technology would also presume the maintenance in perpetuity of a key server to provide access to the public key needed to validate the digital signature.

x   A 'one-way' hash function is a hash function, which is very difficult to reverse. This means that the digest is easy to compute, but it is difficult to determine an input message that will generate a specific digest [sect. 2.4, Christofferssen et al., 1988].

xi The line break tag (i.e. "<LB>") is a milestone tag and specifies a point in the document where a line break occurs. It has no end tag.

xii In fact the transcription process used by the ASEC requires this information to be recorded to aid in the proofreading of the transcription against the original. Lineation and pagination are used as referents back to the original documents.

xiii This "alteration", under the JITM system can only apply to the minimal SGML markup required to uniquely identify the text elements of the transcription. The text of the transcription, by definition, remains that of the original author.

xiv Unfortunately this applies especially to the Macintosh where there is a singular dearth of SGML-based browsers.