When professionals, researchers, or academics in a particular domain need to share information, usually a set of commonly used terms and concepts exist that facilitate communication. This set can be expressed by defining the structures, attributes and contents that describe the domain in a model. The model formally organizes data and provides some meaning, but generic features describing an entire field may not express concepts across specific sub-domains. While field experts are familiar with that area, they are not usually familiar with computer science data model concepts. Adopting a generic model and understanding how to conform to defined structures that are unfamiliar to field experts creates a knowledge gap. We are developing a method that identifies the meaning of generic features within the context of a sub-domain. The proposed solution provides a method to describe what a generic feature represents to the sub-domain in order to support sharing the usage of generic features and create a higher interoperability functionality.
ADOPTING LARGE STANDARDS THROUGH KEYWORDS AND SOCIAL NETWORKING

A Thesis

Submitted to the
Faculty of Miami University
in partial fulfillment of
the requirements for the degree of
Master of Computer Science
Department of Computer Science and Software Engineering

by
Holly Lynn Connor
Miami University
Oxford, Ohio
2011

Advisor ________________________________
Dr. Gerald Gannod

Committee Member ________________________________
Dr. Janet Burge

Committee Member ________________________________
Dr. Laura Mandell

Committee Member ________________________________
Dr. John Timm
# Contents

1 Introduction ...................................................... 1
   1.1 Thesis Statement ........................................... 5
   1.2 Contributions ............................................. 5
   1.3 Organization .............................................. 6

2 Background and Related Work ................................. 7
   2.1 Background .................................................. 7
      2.1.1 XML Interoperability Standards ....................... 8
      2.1.2 Text Encoding Standards ............................... 8
      2.1.3 Clinical Document Standards ......................... 9
   2.2 Related Work ............................................... 10
      2.2.1 Adopting Standards .................................. 10
      2.2.2 Social Tagging and Folksonomy ....................... 11
      2.2.3 Social Networking for Education .................... 12
      2.2.4 TEI By Example ..................................... 12
      2.2.5 Capturing Rationale ................................ 13

3 Approach ......................................................... 16
   3.1 Adoption Challenges ...................................... 17
   3.2 Social Network and Groups ............................... 18
   3.3 Page Content ............................................... 20
   3.4 Keyword Tags .............................................. 23
   3.5 Rationale ................................................... 25
      3.5.1 Rationale Naming Conventions ....................... 26
   3.6 Tag Clouds ................................................ 28
   3.7 Implementation ............................................. 30
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7.1 Searching for Examples</td>
<td>33</td>
</tr>
<tr>
<td>3.7.2 Adding New Examples</td>
<td>34</td>
</tr>
<tr>
<td>3.7.3 Add New Group</td>
<td>36</td>
</tr>
<tr>
<td>4 Evaluation</td>
<td>37</td>
</tr>
<tr>
<td>4.1 User Study Design</td>
<td>37</td>
</tr>
<tr>
<td>4.1.1 Research Questions</td>
<td>38</td>
</tr>
<tr>
<td>4.1.2 Pretest-Posttest Design</td>
<td>38</td>
</tr>
<tr>
<td>4.1.3 Tasks</td>
<td>42</td>
</tr>
<tr>
<td>4.1.4 Pilot Study</td>
<td>43</td>
</tr>
<tr>
<td>4.2 Results</td>
<td>43</td>
</tr>
<tr>
<td>4.2.1 Participant Characteristics</td>
<td>44</td>
</tr>
<tr>
<td>4.2.2 Comprehension Attitude</td>
<td>44</td>
</tr>
<tr>
<td>4.2.3 Task Performance</td>
<td>46</td>
</tr>
<tr>
<td>4.2.4 Experience with TEI encoding</td>
<td>49</td>
</tr>
<tr>
<td>4.2.5 Expectations vs Perception of the Keyword site</td>
<td>51</td>
</tr>
<tr>
<td>4.2.6 Keyword site Feedback</td>
<td>55</td>
</tr>
<tr>
<td>4.3 Threats to Validity</td>
<td>57</td>
</tr>
<tr>
<td>4.3.1 Internal Validity</td>
<td>57</td>
</tr>
<tr>
<td>4.3.2 External Validity</td>
<td>58</td>
</tr>
<tr>
<td>5 Conclusion</td>
<td>60</td>
</tr>
<tr>
<td>5.1 Discussion</td>
<td>60</td>
</tr>
<tr>
<td>5.2 Future Work</td>
<td>61</td>
</tr>
<tr>
<td>Bibliography</td>
<td>63</td>
</tr>
<tr>
<td>Appendix</td>
<td>67</td>
</tr>
<tr>
<td>A User Study Handout</td>
<td>67</td>
</tr>
</tbody>
</table>
List of Tables

3.1 Challenging Encoding Issues ................................................. 17
3.2 Rationale Naming Conventions .............................................. 26

4.1 User Study Research Questions ............................................ 38
4.2 Pretest Sections and Purpose .............................................. 39
4.3 Correlated Pretest and Posttest Statements to Measure Adequacy ............... 40
4.4 Pretest Section - TEI Knowledge .......................................... 45
4.5 Pretest Section - TEI Comprehension .................................... 46
4.6 Pretest Section - Learning to Use TEI ................................... 47
4.7 Summary of Task Performance ............................................. 48
4.8 Posttest Section - Overall Experience of Experiment .......................... 49
4.9 Posttest Section - Experience During Each Task ........................... 51
4.10 Posttest Section - Usefulness of Individual Features ......................... 51
4.11 Correlated Pretest and Posttest Statements ............................... 52
4.12 Pretest Section - Expectations for a TEI social network ..................... 55
4.13 Posttest Section - Perceptions of Keyword Site ........................... 56
4.14 Posttest Section - Usability of Keyword Site ............................. 57
List of Figures

3.1 Poem Encoded using the TEI standard ........................................... 18
3.2 Domain and Sub domain Relationships ............................................ 19
3.3 Example of Keyword site PAGE .................................................. 21
3.4 Page Contents and Relationship between Group and Tags .................... 22
3.5 Using Tags to Represent Multiple Terminologies ................................. 24
3.6 Rationale Naming Conventions on Keyword site ................................. 27
3.7 Tag Cloud on Keyword site .......................................................... 29
3.8 Model of Overall Approach .......................................................... 31
3.9 Front Page of Keyword site .......................................................... 32
3.10 Tools list on Keyword site toolbar ............................................... 32
3.11 Search through Keyword site Groups ........................................... 33
3.12 Create a New Group Page ............................................................ 35
3.13 Rationale Entry on Page .............................................................. 36
4.1 Expectations of a TEI social network and perceptions of the Keyword site . . . . 54
Acknowledgements

First, I would like to thank my advisor, Dr. Gerald Gannod. As an undergraduate, he made me aware of research opportunities and guided me through the approach and implementation of these research projects. He also provided me with advice about continuing education after undergraduate studies and the additional opportunities it presents. As my advisor throughout graduate school, his guidance has made this thesis possible. He also gave me information about an internship opportunity through his former colleague. I was selected to become an intern, which gave me the opportunity to travel out west to California and gained a lifetime of experience. I am extremely grateful for all of his advice and extensive experience he has given to me throughout this thesis and other projects.

I would like to thank Dr. Laura Mandell for her enthusiasm and constant effort to promote student research. She constantly looks for new projects to incorporate English and Computer Science research fields. Her guidance and encouragement made it exciting to embark on projects and develop a tool that promotes learning. I am thankful she was a part of my committee, provided time for her students to participate in the user study and gave her insight about improvements to the research.

I would also like to thank Dr. Janet Burge and Dr. John Timm for serving on my committee. I appreciate the time and effort they spent considering areas of improvement for my thesis.

Thank you to the faculty of the Computer Science and Software Engineering department for the contributions to my education and research. I am thankful for the funding during the two years of graduate school and during my summer research participation. The past six years being a part of this department and Miami University have been one of the best educational experiences I could imagine. The department has so many active faculty members who are interested and willing to take more time and work to advance knowledge, understanding and research.

Finally, thank you to my parents, Jeff and Jill, and fiancé, Eric, for always giving me your love, support and encouragement.
Chapter 1

Introduction

The exchange of information serves as the initial step towards gaining a clearer understanding of some concept or problem. In order to clearly understand digital information, a specific structure needs to be established that can be used to arrange data into meaningful categories. For groups wanting to have a uniform appearance of the content in their documents, a minimal set of requirements or template of a document must be established. This template serves as the example for what future documents should contain. As this practice progresses, many versions of templates and requirements come into existence to accommodate a variety of exceptions. When multiple versions exist, independent groups try to consolidate or transfer information, causing a situation where information can be lost or misinterpreted. This template-based approach can become tedious and inefficient. In order to address this initial problem, users from a field gather to identify common data and terminology that is used by special interest groups. This collaboration leads to the development of standards that formally define and represent the content and organization of data that comes from the domain or sub domains of the field.

Once all constituents have access to the standard, documents adhering to the standard can be exchanged with others in the entire domain regardless of their affiliations within a smaller interest group. These standardized documents provide those in specific interest areas with common
semantics for content in documents. Another major advantage of using a standard is the ability to have documents that are machine-readable. As an example, a standard creates the formality that can express the subject of a numerical value or the context of a simple description. This unifies all information that is present in the specific area and facilitates the storing and transfer of data that is captured or known within the area using automated tools.

Adopting standards to enforce constraints and hierarchy structures can be difficult for domains that encompass many smaller sub domains. For example, consider a standard that has been developed to encode text mainly from the humanities, social science and linguistics. These three groups encapsulate many other types of text such as poetry, letters, etc. A standard has been designed to meet the needs of all those encoding text by defining common features found across these domains (e.g. paragraph, poem line, title, author, text division).

As another example, consider a standard that has been developed for the healthcare field represents data captured when examining and treating a patient. Healthcare contains many special interest groups such as cardiology, pediatrics, dermatology, pharmaceuticals, insurance claims, etc. The standard is applicable to all groups in the field due to common features (e.g. for healthcare record: family history, medications, vital signs) that make up the what the domain expert is interested in documenting.

These features make standardizing information more flexible and can be tailored to the specific cases that occur in the sub groups. This flexibility adjusts to capture information for the domain but can also create ambiguity within sub domains. This ambiguity arises when sub domains have a terminology unique to their special interest that they wish to express. Without customizing the underlying standard and therefore going back to practices that presented the need for a standard, special interest groups need to understand the entire standard and then identify what features are relevant to that sub domain.

This identifies the problem, learning an entire standard in order to identify the specific features relevant to the sub domains is inefficient and increases adoption barriers. These barriers hinder
the movement to unify data within a field and creates expensive redundancies.

This problem can be addressed by taking advantage of the interaction of users from a special interest area or sub domains. Users create documents that contain information specific to that sub domain and generally only use a subset of the features from the standard. Sub domains have collaboration sites (e.g. Wikis) as a way to share what a feature means to others in that sub domain; additional feature interpretations can be added later if the term becomes something frequently used. Although the use of a wiki is collaborative, it is not organized in a way that would allow quick sharing and search.

Creating an environment where users can quickly add their feature interpretation and sub domain usage would reduce the entire standard to the sub domain interest features. Instead of filtering through the large standard, sub domains could focus on this smaller subset of features and find what is needed. This smaller set of features would be further specified for the sub domain and would provide a clearer understanding of how to use the specified features (as opposed to using the generic features).

The Text Encoding Initiative (TEI) standard [26] defines generic features that apply to many types of text; for example, the “div” feature. This feature represents many types of divisions found in a body of text. Using poetry as the sub domain of interest, the “div” feature can be used to identify types of text found in poetry (e.g. poem, story, letter, drama). The ability to viewing the “div” feature only in the context of poetry would allow those adopting the text standard to focus on only those instances. This would reduce the amount of knowledge they need to acquire in order to use the standard for poetry text.

In the healthcare standard example, a generic feature of the standard is given as “Substance Administration”. This structural feature provides the data container that should hold all of the information regarding something a physician can administer. However, what a physician can administers varies greatly, however many sub domains would identify this as a “Medication”. Being able to refer to this more familiar term reduces the time spend looking through documentation and
examples to verify. Generic features interpreted by someone in the sub domain presents features that are classified by the specific interest terminology which is the most familiar to the user. For future users, this reduces the time spent filtering through features that are not relevant to similar sub domain users.

By capturing additional feature semantics or keywords separately, and pairing them with standard features, the standard does not need to be modified (as in the case of the template-based approach) to customize features for sub domains. These keywords being paired with a generic feature from the large standard become the components that are modified for different sub domains. By creating a method and structure to these pairings (as opposed to using a wiki where user’s search through content) these pairings can be evaluated, and the statistics the user is most interested in following can be identified. Using the previous example, a sub domain user could search for “Medication” rather than “Substance Administration” and identify that “Substance Administration” is the generic feature but that it can be used to capture data for a “Medication”.

Through the use of social tagging or social bookmarking, users share bookmarks to online resources with additional metadata beyond the site location [24]. As Storey et al. describe, “the advantages of these social tagging systems is that the user is free to choose any descriptive terms and is not restricted by a preconceived vocabulary, taxonomy or ontology.” Using the approach of social tagging would allow sub domains to create communities for sharing features that have been interpreted by others within that community. These features can be tagged or marked with keywords and presented as the initial terminology a user considers. Using a system of keywords can quickly locate generic features that would be hard to find using documentation and document examples. It is important to note, document examples would only contain generic features in order to preserve the ability to exchange with others.
1.1 Thesis Statement

The creation of a smaller, specific model from a larger model requires extensive knowledge of a standard and domain knowledge that the standard was created to represent. A dual expertise in computing and a specific domain is challenging for a single person to develop and creates a knowledge gap for those with interest in adopting standards but who have a lack of familiarity in either computing or the domain. By building new document models from elements previously identified by more specific domain tags, users can quickly find what generic element meets their needs without requiring them to spend additional time referring to the standards regulation guides.

1.2 Contributions

The contributions of this research are in the areas of Adopting Standards, Social Tagging and Design Rationale. The problem will be addressed by creating a method that combines social tagging and feature rationale from sub-domains to aid in adopting standards. This approach will facilitate the adoption of standards by creating a solid data infrastructure which can express the relationship of sub-domain specific terminologies in relation to the generic domain standard. In addition, by creating a data infrastructure future extensions and customized interfaces can be developed that provide flexibility to end users.

By adding more information to an existing model rather than creating a customized version, users can reference the same model which allows the same validation, or following of a standard correctly and achieving the same understanding, for everyone using the standard. This facilitates the sharing and understanding of feature tagged data by providing domain specific information about a generic feature. Additionally, while the new information is referencing the generic model, it could still be removed and stand on its own for further use. The ability to decouple the two models can allow for multiple uses (creating schemas, uploading keyword/element pairs) without the user recreating the underlying schema and still meeting the standards specifications.
If a user encounters an element that is very specific to their needs and it hasn’t been further defined by others in their field, the user can add theme keywords, terms related to their field and usage rationale to this element for other’s future use. This collaboration can be further extended to large chunks of data that typically have a common set of generic features structuring the data. These features could be identified as a single entity and then annotated for what is domain specific about them. This keyword or annotation can be searched on and the generic features can be returned to the user as one item to incorporate into their document without multiple reference look-ups.

1.3 Organization

The rest of the thesis consists of five chapters. Chapter 2 presents the background and related work of this thesis and the contributions. For Chapter 3 our approach, methodology and implementation are presented. Chapter 4 describes the evaluation of our approach and provides the results of the user study we conducted to answer the research questions used to test our approach. Finally, Chapter 5 summarizes the thesis and gives conclusions and future work for our research.
Chapter 2

Background and Related Work

2.1 Background

In the process of adopting standards, information to be exchanged is commonly formatted using a structured approach. These structured methods enforce that the data is captured in a uniform way that is machine-processable, as well as, human read-able. As XML standards become a part of other specialities not affiliated with information technology, users become less familiar with the technical constraints and practices. This creates issues when these users want to adhere their data to standards but are unable to understand the method of correctly formatting the data. We are interested in addressing this issue by providing a method and environment that allows users to share knowledge that specifically relates to their more specific needs. By centering the focus on elements of the standard that specifically relate to the user, users have a smaller set of information to deal with and therefore are better able to adjust to the standard.

The first part of this chapter presents background work in XML interoperability standards, text encoding standards and clinical document standards. The second part presents related work in adopting standards, social tagging, folksonomy, social networking, *TEI By Example* and capturing rationale.
2.1.1 XML Interoperability Standards

The Extensible Markup Language (XML) is built upon the original approach for marking up data, known as Standard Generalized Markup Language (SGML) [7]. SGML is a technology accepted by the International Organization for Standardization (ISO) for defining generalized markup languages for documents. XML is a standard produced by the World Wide Web Consortium (W3C) to encode documents in a machine-readable form[4]. Markup languages such as these are used to create methods of representing data in electronic forms that are platform-independent. Since XML is a meta-language, or a language that is used to formally describe another language, it can be used to formally define the structure of another language. The extended definitions are typically represented as a Document Type Definition (DTD) or XML Schema Definition (XSD) which verify that XML documents follow the definition correctly. When XML documents conform to a DTD or XSD, anyone else who also has the DTD or XSD can share information and extract the same marked up data.

Using XML to encode documents creates data that is human and machine readable. This approach structures data in a form that is not dependent on a certain platform and can be interpreted by many readily available tools. Typically, data that is machine readable is not a form that humans can read but XML documents consists of tags indicated as a label placed between the indicators “<”, “>” and the end indicator “/>” (e.g. <tag>Tagged</tag>). An XML document that has tags which are defined in familiar terms, such as “tag”, makes data that can be read by humans who are familiar with the format of XML.

2.1.2 Text Encoding Standards

The standard for encoding text for representation in digital form was created by the Text Encoding Initiative (TEI) consortium. TEI standards have adapted over time, first using SGML to express concepts and then later moving into XML based representations that are better able to fully de-
scribe text content. These guidelines are shared mainly throughout humanities, social science and linguistic communities. The use of TEI standards has been adopted by libraries, museums, publishers, and individual scholars [26].

TEI provides guidelines that specify methods for creating machine-readable texts. These guidelines outline how a type of text should be encoded correctly in order to fully capture and describe the contents. Once text has been encoded using TEI, documents can be transformed into webpages to be viewed in online archives, database formats, simple text to analyze text characteristics, displayed using visual interpretation programs, etc.

The TEI community has established various special interest groups and tools to aid in the adoption of TEI. These groups typically use a shared XML schema and XSLT transformations to generate archives and the various other types of storage or presentation methods that exist. In addition, these shared schemas are specifically created to represent the type of text the group is interested in. For examples, a group referred to as Romantic Circles [21] is specifically interested in the Romantic-period of literature. Those in the group creating TEI documents are using schemas and transformations that have been tailored to represent and encode literature specifically found in the Romantic period.

2.1.3 Clinical Document Standards

Clinical document standards are specifically known as the Clinical Document Architecture (CDA). This standard began with Release One (CDA R1) in November 2000 and Release Two (CDA R2) in May 2005, that are approved by the American National Standards Institute (ANSI) [10]. The purpose of CDA is to fully represent any type of clinical information that exists (e.g. text, images, voice recordings, etc.) and be able to address any variability of content across all clinical areas. In addition, the bare requirements of a CDA document is to have content contained inside a structured header and sections of descriptive data, which is human readable. Building on the basic method, advanced adoptions of the standard involve capturing data in the document as coded values from
specific sources. These coded values make the documents computer readable and add additional depth to the semantics, or meaning of the data, in the document.

2.2 Related Work

2.2.1 Adopting Standards

Standards have been used in many areas where requirements or a quality expectation needed to be established. The standard formalizes the necessary features that an item must have in order to be an acceptable design and achieves a similar level of uniformity with other items following the same standard. In order to exchange data, a common format has to be established in order to achieve seamless transfer. Standards are introduced to define a structure and content that strives to make all data uniform. Many standards for the representation of data use XML technologies to represent the requirements and regulations of the standard. Using an XML approach provides many tools for creating documents, as well as extracting data from documents.

The introduction of standards is a step towards being able to exchange data, however, there are many challenges that still exist after the standard is accepted by the community. Within the healthcare field, the adoption of data standards has been slow due to the impression that required data standards do not exist and the lack of documentation explaining how standards are to be implemented. In addition, a competition between multiple standards organizations makes vendor product implementers have to choose among multiple options. One element of Hammond’s proposed solution is to “create a balance between the technology required to make the standards and the clinical domain expertise of defining data and knowledge content of those standards [14].”

Standards adoption also encounters obstacles when the initial cost of adoption outweighs the long-term savings. The cost of purchasing the technology and storing information and the time and cost in order to train users for the new technology concerns organizations considering investing in
such a transition[9]. Healthcare providers are also concerned with the possibility that the standard is reversed or adopting another innovative approach is presented. By committing a large amount of time and money to the adopting standards, the provider can be locked into an ineffective standard. Therefore, providers continue to wait for others to adopt the standard and attempt to avoid buying into a technology that will not become universally accepted.

2.2.2 Social Tagging and Folksonomy

Social tagging systems have become a popular medium where users can actively create, evaluate and distribute information [29]. The advantages of these social tagging systems include allowing the user to choose any descriptive terms and not be restricted to a preconceived vocabulary, taxonomy or ontology [24], the created tags can be used to understand the tagged resources and also to understand a users’ behavior [29]. Social tagging has also been used in source code as a new approach to maintaining code and documentation [24].

Wu and Zhou [29] discussed the use of tagging systems within online communities and the data created in social tagging, or folksonomy. They describe a similar advantage, stating that “tagging gives users the ability to create their own structures for knowledge organization”. In addition, that the tags created are not only related to the resources but also the users. The research they conducted focused on the semantic relatedness with a folksonomy. The relationship between tagging and users is described by Tom Gruber as a three-piece relationship, such as Tagging(resource, tag, tagger) [13]. This function and its input are described as follows: “A tag is a user-defined string, usually a single keyword that is associated to resources in the act of tagging. Tagger is the user who gives the tag to the target resource. Tagging is defined as the process of attaching tags to resources by tagger.” By connecting users and resources through tags, a user can find resources by searching through tags or by looking through the tags of another user who has the same interests.
2.2.3 Social Networking for Education

Social networks are typically associated with connecting to people for communication and sharing personal information. As the use of social networks has become more common, discussions about integrating social networks into formal education have emerged [1, 18]. Several approaches to adapt social networks into formal education uses exist, but these options still encounter many challenges when universities consider to incorporate them. Although these challenges exist, Alexander emphasizes that the “social software’s popularity proves that people very much want to communicate with others online.” In addition, Alexander presents suggestions to move higher education towards social networking services. These suggestions include: “Identifying Successful Architecture” (e.g. add a social factor to course management systems), “Learning from Learning 2.0” (e.g. avoid hindering those wanting to use Web 2.0 technologies) and “Rethinking Literacy and Thinking Critically” (e.g. update lesson content to better inform and engage students to discuss Web 2.0 issues) [1]. Lockyer and Patterson concluded that using social networking sites in formal education had positive learning outcomes and participant experiences. In addition, provide evidence to show that informal learning can be a positive result of social networking [18].

2.2.4 TEI By Example

This project is an online tutorial that describes different stages of adding TEI elements to a document. The project team set out to design a site such as this to address an “urgent need for an on-line TEI course by example which is less generic than the two tutorials published on the TEI website (A Gentle Introduction to XML and TEI Lite: An Introduction to Text Encoding for Interchange) [26] and more general than the 29 guides to local practice on that site [28].”

These tutorials use examples of text that has been encoded using TEI and relate to eight different aspects of text encoding found in the humanities. These eight aspects are: Common Structure and Elements, The TEI Header, Prose, Poetry, Drama, Primary Sources, Critical Editing and Cus-
It also divides the information into modules that reflect different information about TEI. The modules provided are Tutorials, Examples, Tests, Exercises that each have a page referring to one of the aspects mentioned earlier. There is also a Tools module that provides tools that can be used for TEI in general.

The examples throughout the site are real-life examples of text from the given aspect of text markup. These examples are also accompanied with interactive tests and exercises found in the Tests module. These reinforce the examples given and relate the content back to the specific guidelines given in TEI. The exercise module is also available that accompanies the TEI aspect the user is currently learning. The exercises allow you to attempt encoding a given piece of text and validates if the text has been correctly encoded according to the TEI schema.

This project is intended to be relevant to TEI teachers, students learning TEI, and those encoding text and interested in humanities in general. In addition, the examples are also presented with different levels of encoding detail to demonstrate basic and advanced capabilities of encoding using TEI. This serves as a way to introduce examples that are useful for all levels of user experience.

### 2.2.5 Capturing Rationale

In the field of Computer Science, rationale has been used in Human Computer Interaction [17], Software Design [6] and Engineering [20, 15] applications. In these areas design rationale can represent the reasons and justifications behind decisions, the alternatives and tradeoffs associated with other possible decisions and what changes made the decision necessary. Using a design rational system can improve dependency management, collaboration about design, reuse of artifacts, software maintenance, learning and documentation [16]. Nkowocha et al. highlight the benefits of having rationale and use the case of “out- to in- housing of IT Services” as an example. They note that having rationale can serve as a way to observe the change of personnel and explain the decisions of the processes and practices in which they were or will be involved [19]. Further
more, if design rationale is captured with enough detail, it can also be used to identify inconsistent decisions and determine how a change to the design would effect the software system [6].

Although design rationale has all of the mentioned benefits, the ability to achieve thorough design rationale has been difficult. In order to use design rationale and have enough detail to make well-evaluated decisions, extensive details need to be captured and represented in a usable presentation. This level of detail presents those wanting to document decisions with a high learning and discipline curve that requires additional time and effort to develop. The availability of time and budget constraints, the lack of standards and tools, have created challenges during the practice of rationale documentation and therefore, capturing rationale has not been as extensively adopted [19, 25]. In addition, the detail needed to make design rationale robust enough to make inferences requires that all decisions, selected and rejected, are recorded which can be time consuming and expensive [5]. Burge and Brown also state, “the more intrusive the capture process, the more designer resistance will be encountered.”

Addressing the challenges of adopting design rationale practices has led to many approaches. The Decision Representation Language (DRL) was developed by Lee and Lai as a way to express the explicit relationship of decisions made during design processes [17]. The DRL has been an initial starting place for others to create methods that could capture rationale in a standard and detailed format. Tools to capture and represent rationale have been developed to become part of a developer’s workflow and reduce the disconnect felt between developing software and rationale documenting [15, 5].

Representing rationale has evolved into an approach that combines a formal and informal structure. The combination provides computers with a standard format that enables interpretation while still being information that is easy for humans to read. The goal of this hybrid approach is to provide useable data that can be immediately used for computation or documentation. The Rationale-driven Evolution and Management Information System (REMIS) is a tool developed to track, store and display the decisions made during the change of a process model [20]. The basic storage
system of REMIS uses XML and database repositories to represent rationale. By providing an open-format document, separate extensions can be created that can use the data in other applications. In addition, Ocampo and Münch highlighted that using XML provides a “wide choice of technologies to analyze and process the data [20].”
Chapter 3

Approach

In past approaches, adopting standards proved to be challenging, complicated and time-consuming. Those beginning the process of learning a standard typically start with introductory tutorial documents about the general concepts of encoding and work up to the extensive documentation that explains the standard’s specific encoding practices. For example, the TEI Community has a website [26], that provides an introduction to the TEI encoding guidelines, the complete TEI standard documentation, and many other resources to use as reference while learning how to encode with this standard. On the page “Learn the TEI” on the TEI site, a few basic areas are listed as starting points: Encoding Scheme, Modularity and Customization, and Guide to Best Practices. These areas highlight what key concepts a new user must understand in order to continue learning how to use the TEI standard.

The magnitude of reference content that is available is well described by the phrase from the Guide to Best Practices subsection [26],

“In the two-volume print edition of the P4 Guidelines, for example, about one-third of the pages are given to an alphabetical reference list of all TEI elements, but two-thirds are devoted to explanation of concepts, descriptions of proper usage, and examples of tagging”.

16
A high-level of detail is clearly valuable but those who are new to TEI have a hard time distinguishing between essential and advanced concepts, thus creating an adoption barrier.

Our central aim is to develop a way for users to create XML content that adheres to a standard. Specifically, an approach has been developed that allows users to easily find reference material based on context. Once this capability is established, users have reference tags they can search through in order to learn more about how others in their domain have used the standard. This thesis presents a method to combine content tagging, user domains and search capabilities within a social network to create an environment that facilitates learning and reduces adoption barriers. We describe an approach that uses social networks, social tags and decision rationale to create a simplified way of identifying the purpose and context of when to use elements in large standards. For our running example, we are using the TEI standards.

### 3.1 Adoption Challenges

Figure 3.1 depicts a poem encoded using the TEI standard [3]. This TEI document contains the minimal set of XML elements required to encode a type of poem known as a limerick. Even from this simple example, several questions arise. For instance, why is the “lg” (line group) element used? What should I include in the header? The list goes on. In Table 3.1 we have listed six issues that make encoding difficult and the corresponding part of our approach that is designed to address it.

<table>
<thead>
<tr>
<th>Don’t know...</th>
<th>Approach Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) ... who to ask in the community</td>
<td>Social Network</td>
</tr>
<tr>
<td>2) ... the context of what elements to use</td>
<td>Groups</td>
</tr>
<tr>
<td>3) ... how the specific XML syntax looks</td>
<td>Page Content</td>
</tr>
<tr>
<td>4) ... the purpose or use of an element</td>
<td>Keyword Tags</td>
</tr>
<tr>
<td>5) ... the reason why element is needed</td>
<td>Rationale</td>
</tr>
<tr>
<td>6) ... what elements are used most often</td>
<td>Tag Cloud</td>
</tr>
</tbody>
</table>

Table 3.1: Challenging Encoding Issues
The first and second issues involve being able to identify who or what group in the community is appropriate to consult in order to get answers that are the most specific. The third, fourth and fifth issues deal with being able to find desired information and additional options in reference material effectively. The sixth issue concerns the ability to identify any references that has many purposes and then be able to locate all these references quickly. These issues present those new to the standard with many uncertainties that are difficult to clarify using references and tools that cover all guidelines of the standard.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Title</title>
      </titleStmt>
      <publicationStmt>
        <p>Publication Information</p>
      </publicationStmt>
      <sourceDesc>
        <p>Information about the source</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <lg type="limerick" rhyme="abba" n="3">
        <head>Warp Speed, Ms Bright!</head>
        <l>There was a young lady named <rhyme label="a">Bright</rhyme>,</l>
        <l>Who travelled much faster than <rhyme label="a">light</rhyme>.,</l>
        <l>She departed one <rhyme label="b">day</rhyme>,</l>
        <l>In a relative <rhyme label="b">way</rhyme>,</l>
        <l>And returned on the previous <rhyme label="a">night</rhyme>. </l>
      </lg>
    </body>
  </text>
</TEI>
```

Figure 3.1: Poem Encoded using the TEI standard

### 3.2 Social Network and Groups

**Current State.** The current way to find a community member (Issue 1 in Table 3.1), an expert that can assist you in creation of TEI documents, or to identify the context of elements in TEI (Issue 2
in Table 3.1) requires the use of the TEI site or wiki [26, 27]. These sites provide a reference to forums, examples and archives that others in the community have worked on or created, as well as reference materials for how to use different XML elements. A user is faced with sorting through all of the content in order to identify what information relates to their interests. These challenges are being addressed in our approach through the use of a social networks and groups created in the social network.

**Solution.** A social network provides the structure that we need in order represent domain and sub domain relationships, as shown in Figure 3.2. A large domain that contains many smaller sub domains can be represented in a social network by creating special interest user groups. For example, the TEI social network (i.e. domain) could contain special interest groups, such a poetry, drama, or letters (i.e. sub domains). Presenting users with these special interest groups is one of the approaches used within the TEI social network to make search more efficient. When a user is able to choose a group that matches their interest, they drastically reduce the amount of content they need to continue searching through and users can quickly find content submitted by others in that interest group. Once a user finds a group that matches their interest, the content in this group is more likely to contain terms they would search for, which results in finding information quicker and more accurately. In general, the use of groups reduces adoption barriers by organizing the learning content into amounts that are more manageable to look through and are specific to the users’ interests.

![Figure 3.2: Domain and Sub domain Relationships](image-url)
By organizing users in GROUPS in the social network, the infrastructure needed to analyze data in the network and hierarchical organization is enabled. The social networking core functionality is also able to link a user’s contributions with others related to them in the network. Users begin using the Keyword site by creating a user profile and adding any additional information (e.g. about me, brief description, interests, skills) that relates to their experience or use of the TEI standard. This allows the users’ content to be identified and related to their specific interests. Once a user profile has been created, users can join or create any GROUPS that represent domains they are interested in learning more about or are familiar with. These GROUPS are uniquely identified and establish group pages that are the front page to the content relating to that domain’s interest and aids users looking for examples used in that specific context.

**Benefit.** The benefits of using a social network and group capabilities are that users are able to quickly and easily find others in the community and make contact with them. Our approach identifies who would be appropriate to ask by providing profile pages. In addition, the approach provides a mechanism for identifying the XML elements that are most relevant to an affinity group.

### 3.3 Page Content

**Current State.** The current way the user of a standard finds the syntax of an element (Issue 3 in Table 3.1) is to consult the standard’s guidelines or search through wiki content. The guidelines provide extensive details needed to answer these encoding issues while a wiki contains answers provided by past users. Once the user has located the section related to the element of interest, they are required to filter through this information and figure out how the element applies to their document. Our approach addresses this encoding challenge by using PAGES designed to demonstrate the syntax TEI elements.

**Solution.** The second component of our approach addresses how users add content in the social network. PAGES within the social network provide functions that allow users to create, edit and
comment on any content that they post. In addition, PAGES include a field where users can add tags. The use of social tags makes it possible for special interest groups to share and describe content using their specific terminologies while still following the domain standard. We choose to use the functionality of pages and tags to incorporate concepts of social tagging and folksonomy. Figure 3.3 shows an example of a page on the Keyword site, it displays the content’s title, XML syntax, keyword tags and rationale entry area.

![Figure 3.3: Example of Keyword site PAGE](image)

The goal for this part of our approach is to provide a way for users in groups to tag content in order to create a terminology of the group (i.e., folksonomy of the group). Our approach is also able to provide the advantage that Wu and Zhou [29] identified about social tagging (see Section 2.2.2). They highlight the user’s ability to create their own knowledge organization structures which relates to our goal of allowing users to describe content without any structural constraints.
(i.e., naming conventions). Therefore, our approach makes it possible for users to create social tags that are a flexible representation structure.

The PAGES plugin establishes the method of contributing content with keywords and the GROUP plugin can associate pages to the site’s groups. GROUP PAGES direct users to content on the site and that is specifically related to that GROUP. The page contents consist of an XML snippet created using the TEI standard to represent a text from the user’s field of interest. The keywords submitted to the page are used to describe the TEI encoding in terms used in that field of interest or domain. Pairing keyword tags with TEI encodings is how we approach meeting a user’s need to describe generic TEI standard terms in their specific domain terms. Figure 3.4 depicts a model of a group page, the contents of that page (i.e., XML snippet and decision rationale) and the relationship to any tags added to that page. The figure uses UML [22] cardinality notation to indicate that a page is associated with one group and can be associated with zero or more tags.

![Figure 3.4: Page Contents and Relationship between Group and Tags](image)

**Benefit.** The benefit of using PAGES is that three types of reference material (i.e., page content, keyword tags and rationale) are on a page. As previously discussed, our approach uses groups to facilitate the ability to identify special interests. Therefore, by creating pages inside these groups, our approach makes the reference material (i.e., pages) related to the user’s interest easier to find. Our approach integrates the page content (i.e., standard example), keyword tags and rationale into a page of reference material. This provides the user with information about the XML syntax of the element, the purpose or use of an element and why an element is needed in relation to the group context.
3.4 Keyword Tags

**Current State.** The current way to find the purpose or use of an element (Issue 4 in Table 3.1) is also to consult the standard’s guidelines or wiki. The guidelines provide extensive details about the purpose or use of an element and the wiki contains examples provided by past users. The user has to filter through every purpose or use of the element and figure out how it applies to their document. Our approach addresses this encoding challenge by using tags to describe the use or purpose of TEI elements.

**Solution.** Users searching through references without tags typically have to completely read through a web page or perform a browser search in order to identify if the reference content relates to their interest. In contrast, using tags provides a word or phrase that summarizes the content of a page. Presenting the user with keywords tags (also shown in Figure 3.8) is another approach used in the social network to make search more efficient. These tags are links to the detailed content and reduce the content a user is initially presented. The ability to start searching by looking at a word or a few words makes searching for desired content less tedious and time-consuming. In general, the use of tags reduces adoption barriers by creating links using specific keywords that describe the available content.

**TAGS** provide several other relationships that provide additional capabilities. First, the relationship between **TAGS** and **GROUPS** can also be shown by joining a tag, the tag’s page and the page’s group. **TAGS** submitted on group pages are created by group members who will define their posted content in terms related to others in the group. These group tags are more specific which allows users to search through pages quicker and more accurately. Second, the three-piece relationship, Tagging (object, tag, tagger), defined by Gruber [13] can also be expressed through our approach. When a tag is added to a page (i.e., object) by a user (i.e., tagger), this three-piece relationship can be shown by joining the tag, the tag’s page and the page’s owner. As also discussed in Section 3.6, these two relationships can be visually represented through the use of tag clouds.
Our approach of using group defined tags permits the co-existence of specific group terminologies without requiring a change in the underlying generic standard. This preserves the purpose of the standard and allows everyone from a domain to exchange information that adheres to the same underlying standard but allows sub domains to share their specific terminology within that sub domain. Figure 3.5 depicts how different groups can describe the same generic element from the shared generic standard using tags to represent their specific group terminology. For example, the “div” element from the TEI standard. The TEI guidelines state the purpose of this element is for text division in the front, body, or back of text. However, groups have more specific purposes and uses of this element (e.g. Poetry group for poems and sonnets, Drama group for dramas, a Letters group for letters). Therefore, using tags achieves a goal of our approach to enable multiple descriptions of a generic standard without creating customized standards.

**Benefit.** The benefit of using keyword tags is the flexibility and additional detail they provide. Viewing tags in groups establishes the more specific terminology that is used in that special interest area. This allows the domain standard to be explained in sub domain terms that define the purpose and use of an element in the context of that group.

![Figure 3.5: Using Tags to Represent Multiple Terminologies](image)
3.5 Rationale

Current State. The current way to identify why an element is needed (Issue 5 in Table 3.1) is to consult the standard’s guidelines and wiki. The user has to filter through the guidelines for each potential element and figure out which element is the most appropriate use of the standard. Our approach addresses this encoding challenge through rationale to describe why an element is needed or what other elements might be alternative options.

Solution. The third component of our approach integrates how users enter rationale about a page’s content. Rationale created on the Keyword site is held in an area on PAGES called RATIONALE ENTRY. This entry area is an add-on to the basic functions provided by the PAGES plugin. The RATIONALE ENTRY and PAGE content were separated in order to make clean separations between the two information types and saves the data to different storage backends. If future applications want to access and evaluate the data individually, separate storage makes it less complicated to access.

Establishing a formal method of adding tags and rationale provides data that is easier to search and able to record alternative encoding options. While rationale provides details that can’t be expressed using keyword tags alone, separating tags or rationale entry allows new users to adjust to each documentation technique. Separate tag and rationale entry also allows advanced users who are more familiar with the standard to contribute rationale or tags to another user’s content. This provides an extra level of detail that can be used to evaluate element selections made in the future. Additionally, by making the rationale process an optional step, beginners can interact with only tags and the social network in a simple process that doesn’t overwhelm them with tedious formal concepts and requirements.

The goal of our approach is to capture rationale about decisions that were made when creating the particular example (i.e. XML snippet). In our approach we use the Rationale Entry module to provide user’s with the ability to add rationale about the example on a page. The module
allows users to add decisions that were made and alternatives that can be considered regarding the example. A key reason users would become part of the the social network site is to learn about the standard. We want to include this type of information so users have another reference source to promote additional learning. Documenting decisions and alternatives serves as this additional source of information for those using the social network to learn through examples. Therefore, by including the rationale module in our approach, we are including a method that achieves our goal to allow users to capture and view decision rationale about examples.

**Benefit.** The benefit of having rationale is the additional source of reference information to identify why the element is needed. The rationale can provide possible alternative elements without requiring a search through multiple sections of the guidelines. Someone new to a standard can look in one location and get suggestions of related elements without having advanced experience of standard use.

### 3.5.1 Rationale Naming Conventions

In order to establish a common method of documenting and referencing information on the Keyword site Pages and the associated rationale, naming conventions have been defined. This provides structure that can be used in the future to extract rationale for computation and indexing data.

In Table 3.2, the different types of objects that can be referenced are listed. To identify the type of object being referenced, the symbol or identifying brackets are listed next to the type of object that it has been designated to represent.

<table>
<thead>
<tr>
<th>Object</th>
<th>Symbol/Brackets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword Site Tag</td>
<td>#</td>
</tr>
<tr>
<td>XML Elements from Standard</td>
<td>[...]</td>
</tr>
<tr>
<td>XML Attribute from Standard</td>
<td>{...}</td>
</tr>
<tr>
<td>Informal Explanation of Decision</td>
<td>(...)</td>
</tr>
</tbody>
</table>

Table 3.2: Rationale Naming Conventions
To reference *Keyword site tags*, a hash tag # is used before the tag in the rationale entry. *XML elements* are referenced by surrounding the name of the element with square brackets [...]. The traditional XML brackets <..> were not selected since HTML mark-up brackets (which are the same as XML brackets) are automatically removed during network transmission and is also not displayed if a user types content containing that bracket type in the Rationale Entry. *XML attributes* are referenced by surrounding the name of the attribute with curly brackets {...}. *Informal explanations for a decision* are identified by placing parentheses (...) around the informal explanation.

Rationale on the Keyword site is created and displayed through an entry area on Pages. Once a Page has been created, the Rationale Entry area is available for users to create decision rationale about the example XML snippet on the Page. Figure 3.6 displays how rationale appears in the Rationale Entry window. This figure demonstrates rationale that has been created for our running example of TEI encoding. It documents the decisions that the user considered when encoding the given type of text and alternatives that are possibilities to consider when encoding similar text.

![Figure 3.6: Rationale Naming Conventions on Keyword site](image-url)


3.6 Tag Clouds

Current State. The current way to identify what elements are used most often (Issue 6 in Table 3.1) would be to observe the available examples related to the encoder’s interest. In addition, if documents created using the standard have been transformed into database files, these files could be analyzed in order to find common elements. As a result, a user is faced with spending extensive amounts of time to browse documents or search databases. Our approach addresses these challenges by using tag clouds to display tags that have been attached to PAGES on the social network.

Solution. Another component of our approach addresses with how users view tags that are on the site and utilizes the TAG CLOUD plugin of the Elgg software package. Tags that have been created throughout the social network provide simple descriptions as links to content, groups and users that are part of the site. These tags are intended to simplify or summarize the content the tag has been added to and present the user with a way to identify content without randomly searching through any available content.

A tag cloud is a visual display of a weighted list of tag frequencies. The tags appear as words in a box and vary in font size depending on the number of times that word has been submitted as a tag (e.g. the more the word has been used as a tag, the larger the font size of the word in the tag cloud). Figure 3.7 displays a tag cloud from the Keyword site main page, it contains tags that were created to reference different examples of TEI encodings. As the figure shows, words that appear larger are words that were more frequently used on the site as keyword tags to describe an example on Pages.

Throughout the Keyword site, a tag cloud arranges the keywords of the tag in a window and serves as a way to navigate content. The window is included in several areas of the site and displays different subsets of the tags depending on the specific area of the site. The subset of the tags displayed in the cloud are only those that relate to that specific context area (e.g. group home page, user profile page). Tag clouds that are on a group home page are the subset of tags that
have been added to pages only in that group. The tag clouds that a user can add to their profile page or dashboard are the subset of tags the user has created. This type of presentation provides a better understanding of what the standard means in the context of where the user currently is on the Keyword site.

**Benefit.** The benefit of using tag clouds in our approach addresses the challenge of helping users find elements that are used most often. In addition, finding relevant content is enabled through the variety of tags represented in tag clouds. Users who can view many tags at once can quickly filter through the descriptions of the available content and more likely find what they are searching for.

![Tag Cloud on Keyword site](image)

Figure 3.7: Tag Cloud on Keyword site
3.7 Implementation

The social network environment for our example has been created using a software package known as Elgg [11] and is referred to as the Keyword site. In addition to other social networking functions and many plugins for, the site allows submitted content to be described in simple keywords known as tags. We chose this software package because it provides core social networking functionality and many plugins for extending capabilities and features. Elgg has an online forum and developer community to provide reference and discussion areas for those working with Elgg. A rationale module has also been added to pages on the Keyword site. Pages are a place for content to be created and submitted on the site, the rationale module serves as the container on these pages where decisions and alternatives about the content can be created and submitted.

The user’s interaction with our approach is outlined in Figure 3.8. In this figure, the first row of boxes (closest to user) represent groups or tags (e.g. typically viewed in tag clouds) on the Keyword site. These groups or tags are the main source of search terms the user is first presented with when entering the site or locating the home page to begin a search. In the model, these search objects are connected to items displayed in the model as boxes with rounded-edges. These boxes represent pages that contain an example of the standard being used to encode the given type of text or object. Lines connecting the groups and tags to pages represent that there is a link between the search terms and pages. These links are created in several ways: adding the page to a group’s pages, adding the tag to the page or both (i.e. this creates multiple links to the page). The model shows these possible connections by displaying a tag that has been used to describe two pages, a group that contains two pages and a page that is contained in a group’s pages and has a tag added to it.

Our rationale module was developed using the Google Web Toolkit (GWT) [12]. GWT is a software development framework where web based applications using JavaScript can be developed using Java. The toolkit is open-source and developers have used it to create complex browser-based
applications. We chose this software package because some previous work had been done for rationale capture. We reused some of the rationale organization concepts from the previous work and extended those organization concepts through the framework package Ext GWT [23]. This package provides widgets developed using GWT, but provide advanced functionality and appearance that can be tailored to a specific application. There was an online gallery of the package’s available widgets, developer forum and online APIs that provided useful references when using the framework.

Figure 3.9 shows the home page of the Keyword site. The home page displays the newest groups, members and 100 tags that are used throughout the site. The tags shown in the tag cloud on the home page are linked to any content throughout the site where that tag has been added. In group pages and a user dashboard or profile page, tag clouds are concerned with that current group or user.

Once a user has created a username and password from the Keyword site homepage. They have access to login to the site and use the different parts of the site to learn or to contribute content to help others learn. Users without a username and password are also able to view any content on the
site that has been made public.

Once a user is logged in, they generally start their search from the home page and through their user dashboard (i.e. their customized homepage). If users are viewing another part of the site and would like to get to the homepage, they can click the “Keyword” link located on every page of the site. Also, the toolbar located at the top of the screen contains a Tool menu (shown in Figure 3.10) which lists all of the tools on the site that are available to the user.
3.7.1 Searching for Examples

A user that is looking for an example about a certain part of the standard typically starts searching using the tags or the groups on the Keyword site. A search bar is also available, however, when users their search by looking at what is already on the site rather than trying possible search terms, they are more likely to find the information that is related to their interests.

Searching using Groups

To search through groups, a user would browse through the available groups by selecting Groups from the Tool menu on the toolbar. This links to the page that lists groups by latest discussion, newest, popular. Also, groups can be searched by tag if the group isn’t listed in those categories or the user wants to see all groups associated with a tag.

Once the user finds a group they are interested in, they can select that group and go to the group home page. A group home page provides description and tag fields that are similar to a user profile. In addition to the group description and tags, the page also contains discussion, tag cloud and pages that are all related to or about that group. This content on the group home pages is the information users will reference when trying to learn about how the standard is used to represent that group’s special interests.

Figure 3.11: Search through Keyword site Groups
**Searching using Tags**

To search through tags, a user can view 100 of the available tags through the Keyword home page tag cloud. The home page contains the largest view of the tags that are present on the site. If a user is interested in a tag listed in the tag cloud, they click on the tag and are taken to the search results page of the site. The search results present any content that has been tagged with the tag that was clicked. The user also can select from a list of the other items on the Keyword site that can be searched (e.g. users, groups, pages, comments).

If a user searches tags through a tag cloud on a group page, the tags listed there will only return search results that are related to that group. This makes the tag links return results specific to the group. This reduces the scope of the search to only content that is a part of that group’s special interest.

Another way to search using tags is through another user’s profile page. When a user includes a tag cloud on their page and give access to other users to see, they are displaying the tags they have added to content throughout the site. Using this search method allows users to locate other users with similar interests and then view the content they have submitted on the Keyword site by searching with tags listed on their profile.

### 3.7.2 Adding New Examples

To create new examples of XML snippets on the Keyword site, users will typically add the page within a special interest group. If the example applied to general content about the standard, our approach included a general interest group to contain any of these examples and discussion. The user is also able to create a page separate from any group if for some reason they feel this is more appropriate.
Create a Page

To create a group page, the user would click on the "Group pages" option in the left-bar menu. This takes a user to the group pages and has the option for a "New page" in that left-bar menu. The user would add a title and the page entry (i.e. example of the standard) to the page using the given text field and text box. The example can be formatted using basic text editing features provided by the editor at the top of the text box.

![Figure 3.12: Create a New Group Page](image)

Add Tags to the Page

A user is able to add tags in the text field that comes after the page entry text box. To add multiple tags, a user separates each tag word or phrase with a comma. These tags are used to describe the examples in terms that are more specific to that user or the group. These tags allow the group to use terms other than what is used in the generic standard and can provide a simple description of the content in the example.

Add Rationale to the Page

Once a page has been saved, the rationale module is available for a user to document decisions and alternatives. Figure 3.13 shows the rationale entry area on a page after it has been saved. A "New Decision" has been added to the rationale and by right-clicking on the decision, the option to add "New Alternative" or "Remove Selected" is available. To edit a decision or alternative, the user
double-clicks on the row and the field becomes editable. Once the user has finished editing the rationale, they click "Save" and the rationale is updated in the database. Table 3.2 lists the naming conventions that are used to reference tags on the page or elements/attributes in the page’s example and indicate explanations of the decision.

Figure 3.13: Rationale Entry on Page

### 3.7.3 Add New Group

If a user would like to create a new group to represent a special interest for the standard, they can create the group by selecting Groups from the Tool menu on the toolbar. When the user is taken in the groups page, there is an option in the left-bar menu to "Create a new group". This asked the user to enter a description, add tags and enable features of the group. Once the group is created, the user can invite other users of the site, begin adding examples to the site or wait for new members and observe any content that is created by them.
Chapter 4

Evaluation

4.1 User Study Design

We designed and conducted a user study to evaluate the quality of our approach. We identified three areas of interest to focus our study: effectiveness, adequacy and usability of our approach. Our user study was evaluated by using a Pre-Experimental Design known as a Pretest-Posttest Experiment Design [8]. Our experiment was designed using the Pre-Experimental Design approach in order to distinguish that our study does not “meet the scientific standards of experimental designs [2]”. The pretest-posttest experimental design does not utilize a control group which could make it susceptible to another factor outside of the independent variable influencing a change from the time between the pre-test and post-test.

Each participant in the study took a pre-test online questionnaire, given an introduction to our Keyword site, received a list of tasks to perform, and completed a post-test online questionnaire. Both of the questionnaires are organized into sections. In each section, the questions are related to one of our three areas of interest. A participant ID was used to track participant questionnaire responses and their performance during the experiment tasks. The document version of our questionnaires and task assignments is included in Appendix A.
4.1.1 Research Questions

We identified the key points to evaluate during this study by defining three research questions. Our research questions are outlined in Table 4.1. The main research question (RQ1) is concerned with measuring the *effectiveness* of the Keyword site. This examines how much participants are able to learn about how to encode a given type of text. Three sub-research questions were identified to measure the effectiveness of three more specific aspects of the tool. The goal of our other research questions are to collect responses about the design of the Keyword site in comparison to what the participant expected (*adequacy*) and to get feedback about how easy the Keyword site is to use and navigate (*usability*).

- **RQ1**  Can a social network revolving around TEI help users encode text?
- **RQ1.1**  Can the Keyword site be used to identify frequently used TEI elements?
- **RQ1.2**  Can TEI examples on the Keyword site be useful for identifying specific TEI elements?
- **RQ1.3**  Can the Keyword site be useful during encoding tasks?
- **RQ2**  Does the Keyword site meet the expectation for a social network created for TEI?
- **RQ3**  Do users feel the Keyword site is practical to use?

Table 4.1: User Study Research Questions

4.1.2 Pretest-Posttest Design

Participants in a pretest-posttest study design are first given a survey (questionnaire) to evaluate the participant’s initial responses to given questions before the experiment is conducted. The participant is then asked to complete the experiment (i.e. tasks). A second questionnaire is taken after the participant has completed the tasks in the experiment. Other than general background questions, the questions in our questionnaires were written as closed-ended matrix questions. These closed-end questions ask the participants to rate their opinion using a five-point Likert scale (i.e. 1
= “strongly disagree”, 2 = “disagree”, 3 = “neither disagree nor agree”, 4 = “agree”, 5 = “strongly agree”).

Pretest Design

The pretest questionnaire is given to participants before the experiment and consists of five sections. These five sections were designed to gather information regarding the three areas of interest (effectiveness, adequacy and usability). Each of the sections also contains questions that are used to respond to the research questions presented. Table 4.2 lists each of the sections and the intended area it is trying to address. Section 1 of the pretest is a general survey of the participants educational background and job and/or education level. Section 2 has the participant indicate their knowledge level of various aspects of TEI. These provide a basic profile about the type of participants taking part in the user study.

<table>
<thead>
<tr>
<th>Section</th>
<th>Description of Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Educational Background/Level</td>
</tr>
<tr>
<td>2</td>
<td>Knowledge of TEI</td>
</tr>
<tr>
<td>3</td>
<td>TEI Comprehension</td>
</tr>
<tr>
<td>4</td>
<td>Value of Knowing/Using TEI</td>
</tr>
<tr>
<td>5</td>
<td>Expectations for a TEI social network</td>
</tr>
</tbody>
</table>

Table 4.2: Pretest Sections and Purpose

In Section 3 of the questionnaire the participants are asked to rate statements about their comprehension of TEI. If a participant has used other resources to learn about TEI, this could influence their expectations about a social network for learning TEI. To identify any of these possible influences, we also include statements about the participant’s approach to learning TEI encoding practices.

Section 4 attempts to determine the participant’s understanding and attitudes toward TEI. Those who have knowledge about encoding text have varying opinions about the importance and use of
TEI. These opinions could also influence a participant’s motivation towards tools used to learn about TEI. If a user is not familiar or does not agree with using TEI, it is likely they might not understand the issues associated with learning TEI. This section also contains statements for those unfamiliar with TEI, this allows us to identify if their expectations would be appropriate to consider in the results of the user study.

Section 5 provides the participant with a description of the capabilities of a TEI social network. After this description, the participant is given statements (listed in Table 4.3a) about their expectation of such a TEI social network. After introducing the independent variable (i.e. the use of the TEI social network) the participants’ respond to statements about the Keyword site. A section of the posttest was designed to correlate the expectation section of the pretest, these sections serve to assess adequacy. This section of the posttest contains statements (listed in Table 4.3b) to rate the participants’ perception of the Keyword site. The last question in the pretest asks participants for feedback about features they feel would be useful on a TEI social network.

| (a) | Such a social network would help me learn TEI |
| (b) | A social network for TEI does not seem valuable |
| (c) | TEI examples on pages would be a useful encoding reference |
| (d) | Member’s pages of TEI examples would be repetitive and unstructured |
| (e) | Tag searching would make finding correct encoding methods faster |
| (f) | The use of tags for searching is oversimplified and error prone |
| (g) | Pages I create would be helpful to other members with similar interests |

(a) Statements to Measure Expectations of a TEI Social Network

| (a’) | The Keyword site helped me understand how to encode better |
| (b’) | I don’t see the Keyword site providing any additional usefulness |
| (c’) | The Keyword site made it easier for me to locate TEI examples |
| (d’) | The Keyword site creates repetitive examples that are unnecessary |
| (e’) | I found examples faster searching the Keyword site using tag clouds |
| (f’) | Using tags to find Keyword site content is not an effective search method |
| (g’) | The TEI examples, keyword tags and decision rationale will help others |

(b) Statements to Measure Perceptions of Keyword site

Table 4.3: Correlated Pretest and Posttest Statements to Measure Adequacy
**Posttest Design**

The posttest questionnaire is given to participants after the experiment and also consists of five sections. The goal of distributing this questionnaire is to measure the participant’s attitudes and opinions after they have performed the assigned tasks. These tasks give users a reason to interact with the TEI social networking site and gain knowledge about the environment and its functionalities. This experience can then be used to indicate whether they feel the Keyword site implements the proposed features of a TEI social network.

Section 1 of the questionnaire asks participants about their experience while using the Keyword site to perform the tasks. The statements in this section referred to: the difficulty of tasks overall, if the time to complete a task was excessive, whether a task seemed realistic, if it represented a real world purpose, the confidence of the participant while using the Keyword site and their general opinion about whether it was enjoyable to use the site to learn about TEI.

During Section 2 of the questionnaire, the participants are asked about their experience during each task. The information we collect in this section identifies what functionality on the Keyword site was useful or had a purpose for the individual tasks. The results from this section are used to address RQ1.

To measure the participant’s perception of the Keyword site, Section 3 contains statements for the participant to rate that were similar to the questions in Section 5 of the pretest. To give contrast between the pretest and posttest responses, the statements in this posttest section are directly referring to the Keyword site rather than to a TEI social network. Comparing the participants’ responses in this section with the corresponding section in the pretest provides data we can use to identify whether the Keyword site was able to meet the participants’ expectation and address RQ2.

Section 4 asks the participants whether participants used or would find certain features of the Keyword site (e.g. searching tag clouds) useful. Participants rated these features using a scale from 1 = “very useless” to 5 = “very useful”. Concluding the questionnaire questions, section five asks participants to measure the ease of navigation and use (usability) throughout the Keyword site. This
section also identifies if any participants encountered usability issues with certain features which could change the participant’s experience during the experiment and influence their responses. The rating of these statements are used to measure the usability of the tool and address RQ3.

The final page of the questionnaire gives the participant an opportunity to enter any comments or suggestions they might have about the Keyword site or the experiment. These provide feedback for future development on the site and experiment design.

4.1.3 Tasks

The tasks in our experiment were designed to correspond with the sub-research questions of RQ1. These tasks presented three different situations: (1) locate TEI examples that use the same or commonly used TEI elements, (2) identify the specific TEI element used in a particular instance, (3) encode a given piece of text using TEI. These situations require the user to learn or identify the appropriate TEI elements or encoding practice.

The first task (T1) is used to address RQ1.1. It requires that the participant locate what TEI examples are available for a given search tag and to identify what tag is most commonly attached to TEI examples in a given group. During this task, the participants gain general knowledge about what tags are frequently used to describe TEI examples in relation to the entire site and within a group. The second task (T2) addresses RQ1.2 and requires participants to identify specific TEI elements that are a part of the posted TEI examples. The task provides prompts the participant to identify what specific TEI elements are used and examine Groups to locate the TEI examples that are relevant to the given text example. We use the last task (T3) to address RQ1.3. This task presents the participant with a short poem to be encoded. This task breaks down the encoding task into smaller components. First, participants are asked to identify TEI elements that have been used to encode similar examples on the Keyword site. Then, three possible encodings of the poem are presented to the participant and the participant is asked to identify which encoding of the poem is correct.
The three tasks were designed to provide encodings rather than ask the participant to encode given samples of text by hand. This was done to reduce the amount of examples the participant needed to identify considering the minimal experience level expected from the participants. The tasks are also included in Appendix A.

4.1.4 Pilot Study

We conducted a pilot study with our first group of participants to test whether the experiment consisted of reasonable tasks and if the directions were sufficient for the level of experience of the group. With the first group of participants, we found that the first two tasks needed to be less open-ended and provide the participants with more structured responses. In the third task, we identified that the task required more experience than the minimal expectation set for the experiment and was difficult to complete using an online text editor. To adapt the task to a minimal experience participant, we broke the task into smaller components and asked the participants multiple choice questions. In addition, we felt each task also needed more explanation of how the task applied to a real-world instance.

The pilot study also identified some issues with having access to TEI examples on the Keyword site. We adjusted the accessibility settings to give any user the permissions to read the content. We also adjusted some content on the site to clean up TEI examples that contained unclear content and removed additional content added during testing phases.

4.2 Results

This section contains the report of our results obtained from our experiment. Our participants general background, experience and attitude toward TEI comprehension are first described. These characteristics are taken from the statements provided in the pretest). Then, the participants’ performance is reported by pairing together the participant’s performance during the assigned task in
the experiment and their responses given in the posttest. Finally, in order to answer our research questions, we report the results from the participants’ responses in the posttest and reflect on the provided feedback.

4.2.1 Participant Characteristics

19 volunteers with an educational background in English Literature participated in our experiment. All participants were from Miami University. The 19 participants were undergraduate students and their participation was requested through an English Literature course they are currently enrolled in. There were 40 students enrolled in the course, the participants were divided into two groups. The first half of the group were participants in the pilot study and their questionnaires results were used to adjust the experiment for the other half of the participants. The second half of the group were the participants in the experiment and the results collected and presented from these participants consist of those who completed the all parts of the study (e.g. pretest-posttest questionnaires and participated in the experiment).

The participants’ experience regarding TEI guidelines, using TEI enabling technologies and encoding text using TEI were rated on a scale from 1 = “None” to 5 = “Expert”. 16 participants indicated they had no previous experience, while the other 3 participants indicated they had beginner TEI experience with the statements regarding TEI experience.

4.2.2 Comprehension Attitude

Section 3 of the pretest contains statements for participants to rate that are intended to measure their attitude toward comprehension, the usefulness and value of knowing TEI. The results of these two pretest questionnaire sections are described in the rest of this section.
Please rate each statement about your personal experience according to the following scale:
1 - None, 2 - Beginner, 3 - Knowledgeable, 4 - Advanced, 5 - Expert

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning TEI guidelines</td>
<td>84.2% (16)</td>
<td>15.8% (3)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Learning enabling technologies of TEI (e.g. XSD, XML, XSLT)</td>
<td>84.2% (16)</td>
<td>15.8% (3)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Using TEI support resources (e.g. wiki, tools, software)</td>
<td>73.7% (14)</td>
<td>15.8% (3)</td>
<td>10.5% (2)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Developing custom TEI schemas</td>
<td>94.7% (18)</td>
<td>5.3% (1)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Encoding text using TEI</td>
<td>89.5% (17)</td>
<td>10.5% (2)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 4.4: Pretest Section - TEI Knowledge

**Attitude toward learning TEI**

The most common response from participants (on average 15 out of 19) indicated that they didn’t have any previous understanding about TEI or encoding using TEI. The other participants who had some experience learning about TEI indicated that they first, consulted the TEI guidelines online but found themselves lost in the contents. The next most common approach to gaining TEI comprehension was consulting with another TEI user and asking for their guidance.

Overall the participants’ were hearing about TEI concepts for the first time and learning about the purpose of TEI through class discussion and online archive examples.

**Attitude toward TEI usefulness and value**

The participants who had a familiarity with TEI responded to the following statements regarding: the value and usefulness, importance and need of TEI encoding, and taking the additional time to learn TEI. The participants rated statements in this section with values from 1 = “strongly disagree” to 5 = “agree”. The most common response from those who were familiar ranged from 1 to 3, indicating that generally they did not perceive that TEI had much usefulness or value. However, the rating level of 3 was equally common in 2 of the 3 statements in this section, making it difficult...
Please, rate statements on a scale from 1 to 5 to indicate to what extent they apply to you. N/A - never used TEI, 1 - strongly disagree, 2 - disagree, 3 - neither agree or disagree, 4 - agree, 5 - strongly agree. If encoding an unfamiliar type of text, I would:

<table>
<thead>
<tr>
<th>Statement</th>
<th>N/A</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encode the text using the basic set of acceptable elements</td>
<td>84.2% (16)</td>
<td>0.0% (0)</td>
<td>5.3% (1)</td>
<td>0.0% (0)</td>
<td>10.5% (2)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Skip encoding the text and make a note to return later</td>
<td>78.9% (15)</td>
<td>5.3% (1)</td>
<td>0.0% (0)</td>
<td>10.5% (2)</td>
<td>10.5% (2)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Encode using a best guess of elements from a similar type of text</td>
<td>78.9% (15)</td>
<td>0.0% (0)</td>
<td>10.5% (2)</td>
<td>5.3% (1)</td>
<td>5.3% (1)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Look for related documentation in the TEI web site guidelines</td>
<td>73.7% (14)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>5.3% (1)</td>
<td>15.8% (3)</td>
<td>5.3% (1)</td>
<td>19</td>
</tr>
<tr>
<td>Navigate through TEI guidelines but get lost in the extensive contents</td>
<td>68.4% (13)</td>
<td>5.3% (1)</td>
<td>5.3% (1)</td>
<td>15.8% (3)</td>
<td>5.3% (1)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Use tools or software to try to visualize or describe the unfamiliar text</td>
<td>73.7% (14)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>15.8% (3)</td>
<td>15.8% (3)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Search for similar genre archives encoded using TEI</td>
<td>78.9% (15)</td>
<td>0.0% (0)</td>
<td>5.3% (1)</td>
<td>0.0% (0)</td>
<td>15.8% (3)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Search through the TEI wiki pages</td>
<td>78.9% (15)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>5.3% (1)</td>
<td>15.8% (3)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Consult with another TEI user and ask for their guidance</td>
<td>73.7% (14)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>15.8% (3)</td>
<td>10.5% (2)</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 4.5: Pretest Section - TEI Comprehension

to clearly identify a strong opinion in either direction.

Those participants who were not familiar with TEI responded most commonly that they didn’t fully understand what TEI was or how it was used. Although participants were also asked if TEI seems useful and the majority of the responses were rated as a 3 or 4, putting the overall majority of participants on the more agreement side of the rating scale.

Overall, most participants were not familiar with TEI and did not fully understand what it was or how it was used. However, from what has been explained and demonstrated to them most felt that it seems useful.

### 4.2.3 Task Performance

During the experiment, the participants were given an online version of our tasks and access to the Keyword site with a username and password. For T1, we asked participants to identify what type
If you are familiar with TEI, rate each statement on a scale from 1 (strongly disagree) to 5 (strongly agree) to indicate to what extent they apply to you.

<table>
<thead>
<tr>
<th>Statement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowing how to use TEI is valuable and useful knowledge</td>
<td>28.6% (2)</td>
<td>28.6% (2)</td>
<td>14.3% (1)</td>
<td>28.6% (2)</td>
<td>0.0% (0)</td>
<td>7</td>
</tr>
<tr>
<td>Encoding text using TEI is an important and needed practice</td>
<td>28.6% (2)</td>
<td>14.3% (1)</td>
<td>28.6% (2)</td>
<td>14.3% (1)</td>
<td>14.3% (1)</td>
<td>7</td>
</tr>
<tr>
<td>It is worth taking the additional time to learn TEI</td>
<td>28.6% (2)</td>
<td>28.6% (2)</td>
<td>42.9% (3)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>7</td>
</tr>
</tbody>
</table>

If you are NOT familiar with TEI, rate each statement on a scale from 1 (strongly disagree) to 5 (strongly agree) to indicate to what extent they apply to you.

<table>
<thead>
<tr>
<th>Statement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>I don’t fully understand what TEI is or how it is used</td>
<td>11.1% (2)</td>
<td>5.6% (1)</td>
<td>11.1% (2)</td>
<td>27.8% (5)</td>
<td>44.4% (8)</td>
<td>18</td>
</tr>
<tr>
<td>Encoding text using TEI seems to be useful</td>
<td>5.6% (1)</td>
<td>5.6% (1)</td>
<td>38.9% (7)</td>
<td>44.4% (8)</td>
<td>5.6% (1)</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 4.6: Pretest Section - Learning to Use TEI

of content two Keyword site tags were associated with. Most participants (18 out of 19) were able to correctly identify what content the Keyword tags were connected with and which Keyword tag was used most often in a given group’s pages. This task was geared towards getting the participants to utilize the functionality of tag clouds and use them to identify common Keyword tags used and how these tags are related to content.

During T2, we asked participants to find TEI examples related to a given type of TEI text to encode. Once participants were able to locate examples about this type of text, they were able to identify which specific TEI elements were a part of the example. A majority of the participants (13 out of 19) were able to correctly identify the three TEI elements that were present in one example. In addition, a greater majority (15 out of 19) were also able to locate an example that pertained to a specific Keyword site group and identify the specific TEI element that the example was mainly concerned around. The purpose of this task was to have participants be aware of how TEI examples are placed into groups on the Keyword site. This is intended to demonstrate how TEI provides the flexibility to be used for many different types of text. In addition to groups showing types of text, the groups contain examples to show the type of TEI elements used in order to encode that type of text.
T3 required that participants combine the searching techniques of T1 and the observing examples in Keyword groups of T2. This task presented the participants with an example of a poem. The participants were asked to identify which type of TEI elements were to be used when encoding a poem. The slight majority (11 of the 19) of participants were able to correctly identify which TEI element is used to encode the structures of the poem. This task was intended to combine multiple features of the Keyword site to encode something using TEI. The lower rate of success in this task could relate to the additional XML concepts (e.g. understanding XML attributes) required to correctly encode the given poem. This XML concept was not an expected knowledge given our minimal experience required of participants and could have presented information that was unclear to follow in the TEI example. The second part of T3 was to indicate which encoding choice of the poem was correctly encoded using TEI. For this part of the task, two more participants (13 of the 19) were able to identify which TEI encoding choice was done correctly.

<table>
<thead>
<tr>
<th>Task</th>
<th>Performance</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>94.7% (18)</td>
<td>19</td>
</tr>
<tr>
<td>1b</td>
<td>94.7% (18)</td>
<td>19</td>
</tr>
<tr>
<td>2a</td>
<td>68.4% (13)</td>
<td>19</td>
</tr>
<tr>
<td>2b</td>
<td>78.9% (15)</td>
<td>19</td>
</tr>
<tr>
<td>3a.1</td>
<td>57.9% (11)</td>
<td>19</td>
</tr>
<tr>
<td>3a.2</td>
<td>52.6% (10)</td>
<td>19</td>
</tr>
<tr>
<td>3b</td>
<td>68.4% (13)</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 4.7: Summary of Task Performance

Overall, participants felt the task were not very reasonable to perform and felt unsure about how to complete the given tasks. The feedback provided by participants indicated they would have liked more instruction on how to use the Keyword site, as well as, given more instruction about the purpose of the task. Slides were provided as an introduction to the Keyword site, however, a beginning introduction could have been conducted at the beginning of the experiment to give a better understanding of how to use the site.
Please rate each statement on a scale from 1 to 5 to indicate to what extent they apply to you. 1-strongly disagree, 2-disagree, 3-neither agree or disagree, 4-agree, 5-strongly agree.

<table>
<thead>
<tr>
<th>Overall, the tasks were reasonable to perform</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.3%</td>
<td>31.6%</td>
<td>21.1%</td>
<td>36.8%</td>
<td>5.3%</td>
<td>19</td>
</tr>
<tr>
<td>It took longer than expected to complete a task</td>
<td>10.5%</td>
<td>31.6%</td>
<td>36.8%</td>
<td>21.1%</td>
<td>0%</td>
<td>19</td>
</tr>
<tr>
<td>I felt unsure about how to complete some or all of the tasks</td>
<td>10.5%</td>
<td>15.8%</td>
<td>10.5%</td>
<td>15.8%</td>
<td>47.4%</td>
<td>19</td>
</tr>
<tr>
<td>The Keyword site slides were hard to follow</td>
<td>21.1%</td>
<td>5.3%</td>
<td>36.8%</td>
<td>26.3%</td>
<td>10.5%</td>
<td>19</td>
</tr>
<tr>
<td>The tasks were interesting to carry out</td>
<td>5.3%</td>
<td>26.3%</td>
<td>36.8%</td>
<td>21.1%</td>
<td>10.5%</td>
<td>19</td>
</tr>
<tr>
<td>I felt the tasks did not resemble a realistic purpose</td>
<td>10.5%</td>
<td>26.3%</td>
<td>31.6%</td>
<td>31.6%</td>
<td>0%</td>
<td>19</td>
</tr>
<tr>
<td>The tags on the Keyword site were useful</td>
<td>0%</td>
<td>10.5%</td>
<td>21.1%</td>
<td>36.8%</td>
<td>31.6%</td>
<td>19</td>
</tr>
<tr>
<td>I enjoyed using the Keyword site to attempt the experiment</td>
<td>0%</td>
<td>31.6%</td>
<td>26.3%</td>
<td>36.8%</td>
<td>5.3%</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 4.8: Posttest Section - Overall Experience of Experiment

4.2.4 Experience with TEI encoding

We evaluate the effectiveness of the Keyword site and answer our main research question in this section. RQ1 regards whether or not the Keyword site was able to help participants learn about TEI and encode a sample of text. We discuss the three sub-research questions which relate to our main research question and correspond to the three tasks of our experiment.

RQ1.1: Can the Keyword site be used to identify frequently used TEI elements? Participants reported that tag clouds were useful during T1 (a third of the participants also continued to use the tags to search for other pages of examples in T2 and T3) and were able to use the tags in the tag cloud to locate the TEI examples they were searching for quickly. The overall feedback about tags and tag clouds was that they were very useful for locating and describing TEI examples.

We conclude that the Keyword site can be used to identify frequently used TEI elements through the use of tags in tag clouds and group pages.

RQ1.2 Can TEI examples on the Keyword site be useful for identifying specific TEI elements?
Overall, participants reported that observing group pages was the most useful way to find examples during T2 and that the Keyword site as a whole is useful for this task. Once the participants found the group pertaining to the task, they indicated most commonly that viewing examples on the page was very useful, as well as viewing the tags added to the page.

We conclude that creating pages through groups and adding tags to pages on the Keyword site is useful for users looking for examples of how to use specific TEI elements.

**RQ1.3 Can the Keyword site be useful during encoding tasks?** Participants reported that examples of TEI on pages and groups on the Keyword site were the most common approach for learning how to complete T3. In addition, a third of the participants found the Keyword site tags to be useful when locating specific examples of how to encode the given sample of text.

We conclude that the Keyword site is very useful when encoding a given type of text, assuming that the examples of similar content are available on the pages throughout the site.

**Summary.** Overall, all of the participants felt that the Keyword site was a very useful approach for the tasks. Creating groups was rated the lowest in the usefulness section of the responses, however, the tasks had not offered any motivation or purpose to create a group which could effect the participants’ attitudes or experience with this feature of the Keyword site. Although many gave feedback that the purpose of the tasks were unclear and initially understanding how to navigate the site, the performance during all of the tasks had more participants respond correctly than incorrectly. Thinking about these factors, even though participants were sometimes unclear about the tasks and were not familiar the site, more than half of participants were still able to use the Keyword site to successfully and identify the correct choices. Therefore, we can answer **RQ1** positively by stating that the use of the Keyword site can help a new user of TEI during text encoding situations.
Please indicate any tasks for which the following statements apply. Tasks: 1 - Find TEI Using Tag Cloud, 2 - Learn TEI Through Examples, 3a - TEI Elements in a Poem, 3b - Encode Complete Poem

<table>
<thead>
<tr>
<th>The Keyword site is useful for this task</th>
<th>78.9% (15)</th>
<th>47.4% (9)</th>
<th>47.4% (9)</th>
<th>42.1% (8)</th>
<th>10.5% (2)</th>
<th>Total 19</th>
</tr>
</thead>
<tbody>
<tr>
<td>I used the tags to find information for this task</td>
<td>73.7% (14)</td>
<td>31.6% (6)</td>
<td>47.4% (9)</td>
<td>21.1% (4)</td>
<td>5.3% (1)</td>
<td>Total 19</td>
</tr>
<tr>
<td>I started searching through groups to find TEI examples for this task</td>
<td>36.8% (7)</td>
<td>57.9% (11)</td>
<td>52.6% (10)</td>
<td>52.6% (10)</td>
<td>5.3% (1)</td>
<td>Total 19</td>
</tr>
<tr>
<td>I looked for a Group during this task</td>
<td>47.4% (9)</td>
<td>47.4% (9)</td>
<td>42.1% (8)</td>
<td>10.5% (2)</td>
<td>15.8% (3)</td>
<td>Total 19</td>
</tr>
<tr>
<td>I looked for a Page during this task</td>
<td>21.1% (4)</td>
<td>47.4% (9)</td>
<td>52.6% (10)</td>
<td>68.4% (13)</td>
<td>21.1% (4)</td>
<td>Total 19</td>
</tr>
<tr>
<td>I used the search bar during this task</td>
<td>10.5% (2)</td>
<td>15.8% (3)</td>
<td>36.8% (7)</td>
<td>31.6% (6)</td>
<td>42.1% (8)</td>
<td>Total 19</td>
</tr>
<tr>
<td>I saw an example of TEI during this task</td>
<td>31.6% (6)</td>
<td>47.4% (9)</td>
<td>57.9% (11)</td>
<td>73.7% (14)</td>
<td>5.3% (1)</td>
<td>Total 19</td>
</tr>
</tbody>
</table>

Table 4.9: Posttest Section - Experience During Each Task

<table>
<thead>
<tr>
<th>Please rate how useful the features were or could be using the following scale: 1 - very useless, 2 - useless, 3 - neither useless nor useful, 4 - useful, 5 - very useful</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viewing or joining groups representing TEI special interests</td>
<td>5.3% (1)</td>
<td>15.8% (3)</td>
<td>15.8% (3)</td>
<td>52.6% (10)</td>
<td>10.5% (2)</td>
<td>19</td>
</tr>
<tr>
<td>Searching Keyword site using tag clouds</td>
<td>0.0% (0)</td>
<td>5.3% (1)</td>
<td>15.8% (3)</td>
<td>26.3% (5)</td>
<td>52.6% (10)</td>
<td>19</td>
</tr>
<tr>
<td>Creating a group for a new TEI special interest</td>
<td>0.0% (0)</td>
<td>15.8% (3)</td>
<td>47.4% (9)</td>
<td>31.6% (6)</td>
<td>5.3% (1)</td>
<td>19</td>
</tr>
<tr>
<td>Viewing pages of TEI examples</td>
<td>0.0% (0)</td>
<td>5.3% (1)</td>
<td>15.8% (3)</td>
<td>31.6% (6)</td>
<td>47.4% (9)</td>
<td>19</td>
</tr>
<tr>
<td>Observing tags to describe Keyword site content</td>
<td>0.0% (0)</td>
<td>15.8% (3)</td>
<td>15.8% (3)</td>
<td>31.6% (6)</td>
<td>36.8% (7)</td>
<td>19</td>
</tr>
<tr>
<td>Adding tags to further describe pages</td>
<td>0.0% (0)</td>
<td>10.5% (2)</td>
<td>31.6% (6)</td>
<td>42.1% (8)</td>
<td>15.8% (3)</td>
<td>19</td>
</tr>
<tr>
<td>Searching Keyword site using the search bar</td>
<td>0.0% (0)</td>
<td>26.3% (5)</td>
<td>21.1% (4)</td>
<td>36.8% (7)</td>
<td>15.8% (3)</td>
<td>19</td>
</tr>
<tr>
<td>Adding tags to further describe groups</td>
<td>0.0% (0)</td>
<td>21.1% (4)</td>
<td>15.8% (3)</td>
<td>52.6% (10)</td>
<td>10.5% (2)</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 4.10: Posttest Section - Usefulness of Individual Features

### 4.2.5 Expectations vs Perception of the Keyword site

To address the adequacy of our tool and answer **RQ2**, we analyze the expectation vs. the perception of the Keyword site. In the pretest, the participants rated statements about a hypothetical TEI social
network that would provide the functionality of the Keyword site. In the posttest, statements similar to those in the pretest are asked again, but in the posttest the questions refer to the Keyword site rather than a hypothetical TEI social network. By comparing the participants’ responses from these corresponding sections in the pretest and posttest, we can determine whether the perception of the Keyword site matches the participants’ expectation.

The stacked column charts in Figure 4.1 illustrates the results of the expectations from the pretest and the perceptions from the posttest. Each column represents one of the statements regarding a functionality (see Table 4.11) of a TEI social network and the implementation of that functionality on the Keyword site. To indicate the ratings participants’ responded to each statement, the columns consist of up to 5 different colored segments corresponding to each rating of the five-point Likert scale. The statements in (b), (d) and (f) were asked in a negative question format, the charts reflect this alternation of agreement and disagreement responses (i.e. lighter and darker segments of the columns tend to alternate height positions).

Figure 4.1a depicts the ratings regarding the expectations of a TEI social network. The partic-
Participants’ initial reaction to a TEI social network are overall in agreement that a TEI social network would be a helpful approach to learning TEI (a). Most participants felt a TEI social networking has value (b), that the examples would be a useful encoding reference (c) and the examples on such a site wouldn’t be repetitive and unstructured (d). Many participants felt very positive about using tags as a way to search faster (e) and that tags would not be oversimplified and error prone (f). There were an equal amount of participants who felt positive or unsure about whether pages they created would be useful to other users with similar interests (g). In general, participants expectations of a TEI social network for learning were positive.

Figure 4.1b models the ratings participants indicated regarding their perception of the Keyword site. The participants’ were split in agreement about whether the Keyword site helped them encode better (a’), but still agreed that the site made it easier to locate TEI examples (c’). The participants’ agreement about the usefulness of the Keyword site (b’) and disagreement about examples being repetitive and unnecessary (d’) remained the same after using the site. The participants’ increased in agreement about finding examples faster using tag clouds (e’) and felt using tags was an effective search method (f’). The participants also agreed slightly more about the TEI examples, tags and rationale being a help to others (g’). In general, the site didn’t meet expectations of encoding better but it met the expectations of a site that makes it easier to locate examples.

Overall, the participants’ perception about the Keyword site was positive and although participants felt that the Keyword site didn’t help them understand how to encode better, it helped them locate examples faster. Therefore, we can answer RQ2 mostly positive, stating that the site met participants’ expectations of being able to locate examples quickly through the use of tags and tag clouds. Also, that the examples were a useful encoding reference and that would also be helpful to others with similar interests. Reflection on these two corresponding sections provided these thoughts. The first statement in the pretest rated the expectation for learning TEI, while the first statement in the posttest rated the understanding of how to encode. Although these first statements were intended to correspond, this terminology difference could have affected the participant’s in-
Figure 4.1: Expectations of a TEI social network and perceptions of the Keyword site

Interpretation of the question and be responsible for the lower perception after using the tool. This could be argued by highlighting that participants did learn about TEI, as shown in experiment performance. Rather than rating a statement about encoding better, which could be considered a separate expectation, the posttest would have had a better design if the first statement was presented as learning and not encoding.
Note. The handout created for our user study contained statements to evaluate the participant’s expectation of group and rationale functionality in a TEI social network. We removed these statements for our user study with student participants due to most participants’ responding “neither agree or disagree” about these features in the pretest, which seemed to indicate they were unsure about the purpose of the features. This functionality of the Keyword site is intended for users that are very familiar with the site and have more experience with TEI. As a result, throughout the handout we removed tasks and statements pertaining to these features to avoid overwhelming the participants with too many new or inapplicable concepts.

To indicate to what extent they apply to you, please rate each statement on a scale from 1 (strongly disagree) to 5 (strongly agree).

<table>
<thead>
<tr>
<th>Statement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Such a social network would help me learn TEI</td>
<td>5.3% (1)</td>
<td>10.5% (2)</td>
<td>21.1% (4)</td>
<td>57.9% (11)</td>
<td>5.3% (1)</td>
<td>19</td>
</tr>
<tr>
<td>A social network for TEI does not seem valuable</td>
<td>26.3% (5)</td>
<td>47.4% (9)</td>
<td>26.3% (5)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>TEI examples on pages would be a useful encoding reference</td>
<td>5.3% (1)</td>
<td>5.3% (1)</td>
<td>15.8% (3)</td>
<td>57.9% (11)</td>
<td>15.8% (3)</td>
<td>19</td>
</tr>
<tr>
<td>Member’s pages of TEI examples would be repetitive and unstructured</td>
<td>5.3% (1)</td>
<td>63.2% (12)</td>
<td>31.6% (6)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Tag searching would make finding correct encoding methods faster</td>
<td>5.3% (1)</td>
<td>5.3% (1)</td>
<td>21.1% (4)</td>
<td>57.9% (11)</td>
<td>10.5% (2)</td>
<td>19</td>
</tr>
<tr>
<td>The use of tags for searching is oversimplified and error prone</td>
<td>26.3% (5)</td>
<td>47.4% (9)</td>
<td>26.3% (5)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>Pages I create would be helpful to other members with similar interests</td>
<td>5.3% (1)</td>
<td>5.3% (1)</td>
<td>36.8% (7)</td>
<td>42.1% (8)</td>
<td>10.5% (2)</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 4.12: Pretest Section - Expectations for a TEI social network

4.2.6 **Keyword site Feedback**

**Keyword site Usability**

Out of the 19 participants, 9 found that the site was easy to use and 3 “neither agree or disagree” about the site’s ease of use. The layout of the site was a hinderance to quick work for 7 of the 19 participants and 9 felt it was hard to find an example of text they needed. All but 1 participant
The Keyword site helped me understand how to encode better | 21.1% (4) | 21.1% (4) | 15.8% (3) | 21.1% (4) | 21.1% (4) | Total 19

I don’t see the Keyword site providing any additional usefulness | 31.6% (6) | 42.1% (8) | 10.5% (2) | 15.8% (3) | 0.0% (0) | Total 19

The Keyword site made it easier for me to locate TEI examples | 10.5% (2) | 5.3% (1) | 15.8% (3) | 52.6% (10) | 15.8% (3) | Total 19

I wouldn’t use the Keyword site once I understood how to encode with TEI | 10.5% (2) | 47.4% (9) | 26.3% (5) | 5.3% (1) | 10.5% (2) | Total 19

The Keyword site would give me a place to display my special interests using TEI | 5.3% (1) | 10.5% (2) | 42.1% (8) | 31.6% (6) | 10.5% (2) | Total 19

The Keyword site creates repetitive examples that are unnecessary | 21.1% (4) | 36.8% (7) | 42.1% (8) | 0.0% (0) | 0.0% (0) | Total 19

The TEI examples, keyword tags and decision rationale will help others | 0.0% (0) | 15.8% (3) | 21.1% (4) | 42.1% (8) | 21.1% (4) | Total 19

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.1% (4)</td>
<td>21.1% (4)</td>
<td>15.8% (3)</td>
<td>21.1% (4)</td>
<td>21.1% (4)</td>
<td>19</td>
</tr>
<tr>
<td>31.6% (6)</td>
<td>42.1% (8)</td>
<td>10.5% (2)</td>
<td>15.8% (3)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>10.5% (2)</td>
<td>5.3% (1)</td>
<td>15.8% (3)</td>
<td>52.6% (10)</td>
<td>15.8% (3)</td>
<td>19</td>
</tr>
<tr>
<td>10.5% (2)</td>
<td>47.4% (9)</td>
<td>26.3% (5)</td>
<td>5.3% (1)</td>
<td>10.5% (2)</td>
<td>19</td>
</tr>
<tr>
<td>5.3% (1)</td>
<td>10.5% (2)</td>
<td>42.1% (8)</td>
<td>31.6% (6)</td>
<td>10.5% (2)</td>
<td>19</td>
</tr>
<tr>
<td>21.1% (4)</td>
<td>36.8% (7)</td>
<td>42.1% (8)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
<tr>
<td>0.0% (0)</td>
<td>15.8% (3)</td>
<td>21.1% (4)</td>
<td>42.1% (8)</td>
<td>21.1% (4)</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 4.13: Posttest Section - Perceptions of Keyword Site

said they would be able to adjust to using the Keyword site as an encoding aid and 9 felt that the Keyword site was a solution to a TEI social network.

One participant who responded the the site was easy to use, also commented, "I think the site is a good learning tool because of it’s interactive nature". Another participant noted that the site was great and very helpful, in addition, "I used to think learning program code was boring but the Keyword site made it fun and easy". Other participants commented about navigating the site (as mentioned earlier, slides were included to address this however most had not observed them and having an in-class introduction would have ensured some familiarity) and that there needed to be more examples on the site. These issues would be addressed when users have more time to get familiar with the site and as other users contribute new examples of TEI.

We can address the usability of the Keyword site and answer RQ3. Based on these participant responses, we conclude positively, assuming that users would step through a basic introduction to the Keyword site.
Table 4.14: Posttest Section - Usability of Keyword Site

<table>
<thead>
<tr>
<th>Statement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>I found the Keyword site easy to use</td>
<td>10.5% (2)</td>
<td>26.3% (5)</td>
<td>15.8% (3)</td>
<td><strong>36.8% (7)</strong></td>
<td>10.5% (2)</td>
<td>19</td>
</tr>
<tr>
<td>I would be able to adjust to regularly using the Keyword site for encoding aid</td>
<td>5.3% (1)</td>
<td>0.0% (0)</td>
<td><strong>47.4% (9)</strong></td>
<td>42.1% (8)</td>
<td>5.3% (1)</td>
<td>19</td>
</tr>
<tr>
<td>I thought it was hard to find the page of examples I needed to encode my text</td>
<td>15.8% (3)</td>
<td>21.1% (4)</td>
<td>15.8% (3)</td>
<td><strong>31.6% (6)</strong></td>
<td>15.8% (3)</td>
<td>19</td>
</tr>
<tr>
<td>The layout of the Keyword site hindered my ability to work quickly</td>
<td>5.3% (1)</td>
<td><strong>47.4% (9)</strong></td>
<td>10.5% (2)</td>
<td>21.1% (4)</td>
<td>15.8% (3)</td>
<td>19</td>
</tr>
<tr>
<td>A TEI social network seems promising, but the Keyword site isn’t the solution</td>
<td>15.8% (3)</td>
<td>31.6% (6)</td>
<td><strong>36.8% (7)</strong></td>
<td>15.8% (3)</td>
<td>0.0% (0)</td>
<td>19</td>
</tr>
</tbody>
</table>

4.3 Threats to Validity

4.3.1 Internal Validity

Participants

To evaluate the knowledge participants felt they had before given the tasks, we provided statements for them to rate regarding their familiarity with TEI and other topics related to the use of TEI. In addition, as there is some controversy within the literature field about the importance and use of TEI, which we felt could also affect a user’s opinions about a social network created for TEI. Therefore, we asked participants to rate statements about the importance and usefulness of TEI.

Questionnaires and Tasks

The use of a 5-point Likert scale could have influenced participants to answer following a common pattern of rating responses. To reduce the possibility of statements being answered in a pattern, we alternated statements into a positive and negative questioning design. To serve as an introduction to the Keyword site, a set of slides were provided that explained the layout of the site in a high level approach. These were intended to give every user the same basic knowledge about how to navigate the site and avoid in-person assistance which could make the tasks too easy to answer and
influence the participant’s effectiveness and usability responses. In order to evaluate performance uniformly, the tasks were designed so that the participant’s goal ended at finding examples of TEI that we had added to the site. This could influence the participant’s attitude towards the adequacy of the site by making it appear to contain all possible search content, when in reality, the tasks were designed to lead to examples we knew existed.

In addition, the design of assigned tasks to meet the required minimal knowledge could influence the perception of the effectiveness and adequacy as well. Participants with minimal knowledge about TEI would not have the experience needed to create their own examples, groups or tags. This reduces the range of functionality the participant can experience and could limit their expectation about the Keyword site’s usefulness or importance.

**Pilot Study**

The first experimental run uncovered many issues with the level of knowledge the experiment initially required. The initial run contained many open-end tasks that overwhelmed most participants and resulted in most task assignments containing no text encoding. The experiment was modified to be conducted with participants who had no prior knowledge about TEI or encoding. These modifications re-worked questions into multiple choice options and provided the participant with more instructions about the TEI situation. While the modifications made the tasks more feasible for participants to complete, it could have limited the usefulness of the Keyword site by reducing the learning material.

**4.3.2 External Validity**

**Participants**

The participants in our pre-experimental study were from the university student body. Although generally users on the Keyword site would have beginning knowledge and interest with TEI,
we modified our pre-experimental questionnaires and tasks to evaluate participants with no prior knowledge of TEI. To give participants basic information and purpose of the Keyword site, we gave a brief introduction to TEI and what it purpose to serves in the literature field. However, having no prior knowledge or interaction with TEI could affect overall results. The motivation of participants might have also affected the response we received about the usefulness due to participants being unfamiliar with why someone would want to know how to encode text. Also, if participants have had no prior exposure to guidelines or other tools used to learn TEI, they might not realize what the Keyword site provides in comparison to past learning approaches or tools.

Tasks

Our tasks were designed to demonstrate some basic encoding issues that a person encoding text would consider. The tasks had participants use the Keyword site to learn about different elements of TEI and were designed to rely on features of the Keyword site that would be useful for the task (e.g. identify tag frequency which tag clouds are useful for).
Chapter 5

Conclusion

5.1 Discussion

In our research, we developed the Keyword site, which allows users to quickly find examples of a standard that relate to their specific interests. Our method uses social tagging to make it possible for a special interest group to describe a shared generic standard using terms that are specific to their interest area. These tags creating through social tagging are displayed using a navigation view known as a tag cloud. The use of tag clouds gives users an overview of the content that is available on the site and facilitates the ability to search for content relevant to the user’s specific interest or need. Our research also uses the documentation of decision rationale to further describe examples added to pages on the social network. We incorporated the ability to capture decision rationale to give users another way to learn about the standard and the considerations that were made for the example the rationale is associated with.

The standard we selected to use throughout our approach and evaluation was the Text Encoding Initiative (TEI). The Keyword site contains examples that were created using TEI. Groups on the site are special interest areas related to the literature field. These groups served to demonstrate for how TEI is used differently depending on the type of text to be encoded.
Our hypothesis was that a social network created to provide examples of a standard would reduce adoption barriers and facilitate learning about the use of the standard in special interest areas. We tested our hypothesis through a user study by answering research questions we designed that measure the effectiveness, adequacy and usability of the Keyword site. We conducted our user study using a pretest-posttest pre-experimental design. The first part of our user study had participants take a pretest questionnaire about their education background, TEI experience and expectations of a social network used to learn TEI. Then, participants were asked to complete a set of given tasks that require knowledge about TEI by using the Keyword site to learn. Finally, the participants were asked to complete a posttest about their perception of the Keyword site and usefulness of various features of the site. Our results of the pretest and posttest data showed that some users found the Keyword site difficult to navigate and find information. However, that once an example was found it was a useful reference to learn about TEI. Overall, the participants’ expectation of the use of tags was met with in the implementation of them on the Keyword site and participants felt they were very useful. In addition, although participants were not completely motivated to learn about TEI, their performance during the tasks concluded that our site was able to help the participants complete the tasks successfully a majority of the time.

5.2 Future Work

An area of future work is to identify additional features that would improve the Keyword site. Participants provided feedback at the end of the posttest, it would be beneficial to evaluate the common issues mentioned in this feedback. Once the feedback is evaluated, creating a list of the most useful or important features can be created as a list of upcoming features to provide.

Rationale has been shown to serve many purposes in other fields, however, in our research we were unable to find participants for our user study who had enough experience needed to use the rationale functionality. An area of future work is to conduct a user study with participants who
have more than beginner knowledge of TEI (or another standard if a separate social network is created). If the results of this study regarding rationale on the site are positive, this would lead to other future work to identify additional uses of rationale and implement these features in the rationale module.

Finally, our rationale module currently only provides documentation functionality. When enough rationale is available, rationale can be used to verify that the decisions made were appropriate. This presents an area of future work to make additional use of the rationale submitted on the Keyword site. It would be useful to create a system that was able to suggest related material or alternative options to consider when using the standard of interest.
Bibliography


Appendix A

User Study Handout
Keyword Site Experiment

Introduction

Standards for encoding text and other data types were created to structure the data from various domains into machine-readable representations. These standards are flexible to accommodate the range of data possibilities but also adhere to strict guidelines in order to establish machine-readable content. While using standards has many advantages, those learning how to adopt the standard are often faced with a complex and time-consuming task.

Our approach to address these adoption barriers deals with users collaborating and sharing examples created using the standard within a social network. This social network is referred to as the Keyword Site. The Keyword site provides an environment where members can share their examples or browse through other’s examples of text encoded using TEI.

This experiment concerns a member using the Keyword site to learn about correctly encoding data using the given standard.

The given standard in this experiment is the Text Encoding Initiative (TEI). From the TEI site:

“The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics.”

You will become a member of the Keyword site and use it to learn more about using TEI. You will be asked to perform 3 tasks. You are given 20 minutes to perform each of the 3 tasks.

We kindly ask you:
• to write down your answers in a legible way;
• not to consult any other participant during the experiment;
• to perform the tasks in the specified order;
• to write down the current time before starting each task and after completing all the tasks;
• not to return to earlier tasks because it affects the experiment;
• to fill in the required information for each task and give additional information, if requested.

There is a questionnaire before and after the experiment.

Thank you for your participation in this experiment!

Holly Connor and Gerald Gannod
Pretest Questionnaire

General Survey
In this short survey a number of questions regarding your experience and attitude towards TEI comprehension will be asked to get an impression of your skills and expectations.

1 The first part of the questionnaire is about your knowledge and experience with TEI. Please answer these questions about your personal background. These answers will be kept private and only serve to put other responses in context.

What is your education background (e.g. field of study)? .......................................................... .......................................................... .......................................................... ..........................................................

What is your current job/education level? ......................................................................................... .........................................................................................

2 This part of the questionnaire contains questions about your knowledge of TEI. Please rate each statement about your personal experience according to the following scale: 1-None, 2-Beginner, 3-Knowledgeable, 4-Advanced, 5 Expert

| Learning TEI guidelines                                                                 | 1 | 2 | 3 | 4 | 5 |
| Learning enabling technologies of TEI (e.g. XSD, XML, XSLT)                           |   |   |   |   |   |
| Using TEI support resources (e.g. wiki, tools, software)                             |   |   |   |   |   |
| Developing custom TEI schemas                                                      |   |   |   |   |   |
| Encoding text using TEI                                                            |   |   |   |   |   |
**Pretest Questionnaire**

Some statements regarding TEI comprehension activities are shown below. Please, rate each statement on a scale from 1 to 5 to indicate to what extent they apply to you. 
1-strongly disagree, 2-disagree, 3-neither agree or disagree, 4-agree, 5-strongly agree.

**When encoding an unfamiliar type of text, I:**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encode the text using the basic set of acceptable elements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skip encoding the text and make a note to return later</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Encode using my best guess using elements from a similar type of text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Look for related documentation in the TEI web site guidelines</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Navigate through TEI guidelines but get lost in the extensive contents</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use tools or software to try to visualize or describe the unfamiliar text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search for similar genre archives encoded using TEI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search through the TEI wiki pages</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consult with another TEI user and ask for their guidance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TEI allow users to encode text which creates machine-readable documents. Documents that follow a defined standard can then be easily exchanged with others using the standard, the encoding gives content identification and can be transformed into the future data formats. If you are familiar with TEI, rate each statement on a scale from 1 (strongly disagree) to 5 (strongly agree) to indicate to what extent they apply to you.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowing how to use TEI is valuable and useful knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Encoding text using TEI is an important and needed practice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is worth taking the additional time to learn TEI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If you were not familiar with TEI, please rate the following statements on a scale from 1 (strongly disagree) to 5 (strongly agree) to indicate to what extent they apply to you.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>I don't fully understand what TEI is or how it is used</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Encoding text using TEI seems to be useful</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The final part of this survey is about your expectations for a TEI social network. To indicate to what extent they apply to you, please rate each statement on a scale from 1 (strongly disagree) to 5 (strongly agree).

“With a TEI social network, members should be able to add pages that contain examples of text that has been encoding using TEI. These pages also have a field for social tags to be added, these provide a short description of the page. Tags are viewed throughout the site through tag clouds. Tag clouds are a way to visualize the keywords used for the tags and the frequency of the tags on the site. Members can also create or join groups to represent their special interests and knowledge. Additionally, members can add rationale to each page to document the decisions they made when creating their posted TEI example. These decisions also allow alternatives to be attached so others can view what options are available for their needs.”

<table>
<thead>
<tr>
<th>Statement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Such a social network would help me learn TEI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A social network for TEI does not seem valuable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEI examples on pages would be a useful encoding reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Member’s pages of TEI examples would be repetitive and unstructured</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tag searching would make finding correct encoding methods faster</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The use of tags for searching is oversimplified and error prone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pages I create would be helpful to other members with similar interests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Which other features would you find useful in a TEI social network?
Experiment

To get familiar with the Keyword site layout and options, please start with the included slides.

Once you have an idea of the layout of the Keyword site, please continue to the tasks.

Tasks

You can write your username and password details here if you want a reminder location:
Username: ........................................................................................................................................
Password: ........................................................................................................................................

Starting Time: ________________________
Task 1 – Find TEI Using Tag Cloud

Tag Clouds are used throughout the Keyword site to show the tags that have been added to TEI examples. These tags describe content in terms that are more commonly used to describe text.

These can lead to the formal TEI representation without the user being aware of those specific terms.

The Tag Cloud found on the homepage (found by clicking on "Keyword" at any time) lists all the tags, Group Tag Clouds list tags found in that Group’s pages and Tag Clouds on profile pages list tags that user has added.

Answer the following questions using Tag Clouds:

What pages have the tag "line group"? ……………………………………………………………………………………………
…………………………………………………………………………………………………………………………………………………………
…………………………………………………………………………………………………………………………………………………………

Which tag is the most frequent on the Drama Group pages? …………………………………………………
…………………………………………………………………………………………………………………………………………………………
…………………………………………………………………………………………………………………………………………………………

Current Time: ___________________
Task 2 – Learn TEI Through Examples

The Keyword Site organizes TEI examples by Groups. These examples show which elements (e.g. `<oneElement>...</oneElement>`) are used to encode that type of text using TEI.

Elements sometimes contain other elements, for example:

`<outsideElement>`
`<insideElement>...</insideElement>`
...
`</outsideElement>`

Elements sometimes contain text, for example:

`<name>Your Name Here</name>`

The "..." used throughout the Keyword generalizes the TEI examples and can indicate that other elements or text could be there.

Use the examples on the site to answer these questions:

List which elements (e.g. `<oneElement>...</oneElement>`) are found inside the minimal `<teiHeader>...</teiHeader>`:
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
.........................................................................................................................
..................................................................................
William Wordsworth (1770-1850) - A Slumber did my Spirit Seal

1. A slumber did my spirit seal;
2. I had no human fears:
3. She seemed a thing that could not feel
4. The touch of earthly years.
5. No motion has she now, no force;
6. She neither hears nor sees;
7. Rolled round in earth's diurnal course,
8. With rocks, and stones, and trees.

<TEI>
<teiHeader>...
text>
<body>

A slumber did my spirit seal;
I had no human fears:
She seemed a thing that could not feel
The touch of earthly years.

No motion has she now, no force;
She neither hears nor sees;
Rolled round in earth's diurnal course,
With rocks, and stones, and trees.

</body>
</text>
</TEI>
For the following task, determine how to encode the poem below using the Keyword site.

The basic TEI elements are included around the poem located inside the <body>...</body>.

Your task is to identify which encoding example is correct.

What TEI element indicates the text is encoded as a poem?

A. <poem type="stanza">...</poem>
B. <div type="poem">...</div>
C. <stanza type="poem">...</stanza>
D. <poem type="poetry">...</poem>

How should a stanza in the poem be encoded?

A. <stanza>...</stanza>
B. <lg type="stanza">...</lg>
C. <div type="stanza">...</div>
D. <stanza type="lg">...</stanza>
Task 3b – Encode Complete Poem Using TEI

The following are choices of a correct way to encode the given poem.

A. `<div type="poem">
   <lg type="stanza">
     <l>A slumber did my spirit seal,</l>
     <l>I had no human fears;</l>
     <l>She seemed a thing that could not feel</l>
     <l>The touch of earthly years.</l>
   </lg>
   <lg type="stanza">
     <l>No motion has she now, no force,</l>
     <l>She neither hears nor sees;</l>
     <l>Roll'd round in earth's diurnal course</l>
     <l>With rocks and stones and trees.</l>
   </lg>
</div>`

B. `<poem type="div">
   <lg type="stanza">
     <l>A slumber did my spirit seal,</l>
     <l>I had no human fears;</l>
     <l>She seemed a thing that could not feel</l>
     <l>The touch of earthly years.</l>
     <l>No motion has she now, no force,</l>
     <l>She neither hears nor sees;</l>
     <l>Roll'd round in earth's diurnal course</l>
     <l>With rocks and stones and trees.</l>
   </lg>
</poem>`

C. `<div type="poem">
   <stanza type="lg">
     <li>A slumber did my spirit seal,</li>
     <li>I had no human fears;</li>
     <li>She seemed a thing that could not feel</li>
     <li>The touch of earthly years.</li>
     <li>No motion has she now, no force,</li>
     <li>She neither hears nor sees;</li>
     <li>Roll'd round in earth's diurnal course</li>
     <li>With rocks and stones and trees.</li>
   </stanza>
</div>`
For the following task, determine how to encode the poem below using the Keyword site.

The basic TEI elements are included around the poem located inside the `<body>...</body>.

Your task is to identify which encoding example is correct.

Which of the above choices is the correct way to encode the given poem? 

Ending Time: ___________________
### Experiment Evaluation
Thanks for completing the tasks! To get a summary of your experiences with the Keyword site and to allow you to give your comments, please respond to the following questions.

1. The first part is about your overall experience in performing the experiment. Please rate each statement on a scale from 1 to 5 to indicate to what extent they apply to you.
   1-strongly disagree, 2-disagree, 3-neither agree or disagree, 4-agree, 5-strongly agree.

<table>
<thead>
<tr>
<th>Statement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall, the tasks were reasonable to perform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I felt there wasn’t enough time for each task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I felt unsure about how to complete some or all of the tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Keyword site slides were hard to follow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The tasks were interesting to carry out</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I felt the tasks did not resemble a realistic purpose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The tags on the Keyword site were useful</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I enjoyed using the Keyword site to attempt the experiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. This part is about your experience with the Keyword site for each of the tasks. Please indicate every task for which the following statements apply.

<table>
<thead>
<tr>
<th>Statement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Keyword site is useful for this task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I created a page and tags while performing this task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I created rationale while performing this task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I used the tags already added to find related TEI examples for this task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During this task, I used the rationale to find an alternative encoding option</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I started my search through groups to find specific TEI examples for this task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Before the experiment you were asked about your expectations for the Keyword site. Please rate each statement on a scale from 1 (strongly disagree) to 5 (strongly agree) to indicate to what extent they apply to you.

1. The Keyword site helped me understand how to encode better.
2. I don’t see the Keyword site providing any additional usefulness.
3. The Keyword site made it easier for me to locate TEI examples.
4. The Keyword site creates repetitive examples that are unnecessary.
5. I found examples faster searching the Keyword site using tag clouds.
6. Using tags to find Keyword site content is not an effective search method.
7. The TEI examples, keyword tags and decision rationale will help others.

This part is about the usefulness of individual features on the Keyword site. Please rate how useful the features were on a scale from 1 to 5:
1 - very useless, 2 - useless, 3 - neither useless nor useful, 4 - useful, 5 - very useful.

1. Joining or creating groups representing TEI special interests.
2. Creating pages to post TEI examples.
3. Adding keyword tags to TEI example pages.
4. Documenting decisions using Rationale Entry.
5. Searching Keyword site content using tags.

Statements in this part relate to the usability of the Keyword site. Please rate the following statements on a scale from 1 (strongly disagree) to 5 (strongly agree).

1. I found the Keyword site easy to use
2. I would be able to adjust to regularly using the Keyword site for encoding aid
3. I thought it was hard to find the page of examples I needed to encode my text
4. The layout of the Keyword site hindered my ability to work quickly
5. A TEI social network seems promising, but the Keyword site isn’t the solution
Posttest Questionnaire

Enter comments and/or suggestions you may have about the **Keyword site**. These will be useful for future improvements.

Enter comments and/or suggestions you may have about the **experiment**. These will be useful for future studies.