Student readers as authors: Analyzing markings made on e-documents in private or shared reading environments

Asim Qayyum
University of Toronto

ABSTRACT

The main purpose of this study was to examine participating readers' text markings when they interacted with electronic documents during an active reading process. The readings took place in a private setting, where the readers read e-document for self-use, and in a document sharing setting where each document was shared among a group of students. Based on the results from descriptive and statistical tests, a user-marking taxonomy was derived, which led to the development of a framework for marking ontology. The proposed ontology can then be utilized by web searching agents to enhance the information retrieval capabilities of users working with digital library resources in a semantic web environment.

1. INTRODUCTION

Minimal research has been conducted on text marking in electronic documents *(e-documents)* when compared to the research that has been conducted on paper documents markings. The available research on e-document markings focuses mainly on only a few aspects of markings made in the margins. This is in sharp contrast to a substantial number of studies focusing on markings made on paper documents. The paper markings have been investigated by literary and historical scholars (Coleridge 1985; Jackson 2001; Tribble 1997; Stoddard 1985), by scholars in the field of educational psychology, (Nist and Kirby 1989; Idstein and Jenkins 1972; Fowler and Barker 1974), by researchers focusing on collaborative authoring settings (Neuwirth et al. 1990; Adler et al. 1998), and finally, by computer scientists (Marshall 1997; Wolfe and Neuwirth 2001; Marshall and Brush 2002).

Barring one study (Marshall and Brush 2002) that very briefly attempts to characterize, quantify, and analyze online markings, no other relevant research exists that directly goes towards explaining how users interact with electronic annotating applications. Similarly, there are no investigations into document markings in a document sharing setting when a reader is undertaking active reading[1]. The investigation of active reading is essential as we head towards an age of digital libraries; the cyber users will now be able to browse through electronic documents, crossing the boundaries between reading and editing. As a result, it is important to understand how a successful and user-friendly digital library interface can facilitate this new role of library patrons.

The over-riding assumption in previous studies and system development efforts has been that when interacting with e-applications, users behave in the same manner as they do when interacting with paper docs. Such an assumption is not necessarily correct and other than hardware limitations, perhaps it is one of the reasons why so few users are actively using electronic marking applications. Thus, little attempt has been made in the past to develop a taxonomy of online marking patterns in the Human-Document Interaction (HDI) process, and whatever exists in the literature has been derived from studies of paper-based markings. We intend to fill this knowledge gap by performing a study of how users read and mark documents in a private and a document sharing environment.

[1] Indicates reading activity where the reader acts an author, and engages in writing on the text while reading, or takes notes in a separate notebook.

The existing research on annotations made in collaborative authoring settings focuses mainly on comments or notes passed around during the co-authoring of manuscripts (Neuwirth et al. 1990). Hard copy annotations usually prevail in this process as papers are passed among the group members. Within this collaborative authoring process, the most significant activity concerning documents are reading, followed closely by annotating, then collaborating and finally authoring (Brush et al. 2001). Thus, when communities of practice (Wenger 1998) form to tackle a common need, the common needs can be identified in shared documents as these communities work to classify and create their own metadata. A taxonomy of user markings will allow the development of a uniform schema that should be practical and workable for these ad hoc workgroups. Moreover, a common vocabulary can be created using an ontology, allowing group members working in a common domain to share information.

Lastly, note the use of the word *interacting* rather than *reading*. This is because interacting is an all-encompassing term, similar to the expression "active reading", which we will use to cover reading, marking and navigating within an e-document. We believe that given a choice, readers of an e-document will undertake all of the above actions, just as they do on paper.

2. Methods and approaches

2.1 Research questions & hypotheses

The purpose of this study was to explore the interaction of graduate students with electronic reading applications, which also function as electronic writing platforms. Our goal was to understand the user reading and markings activities for sense-making purposes. To achieve this research purpose, reading and marking activities were catalogued while the electronic *(e-)* documents were being read, and possibly marked. This data then were analyzed statistically to seek, examine, discover, identify, and categorize the significant interactions, themes, and sense making patterns.

The investigation was carried out in two types of electronic environments. In the first case, readers read the electronic documents individually, while in the second case a document was shared by passing it to another user. Graduate students were chosen as the target population in this study and the choice was influenced both by convenience and because graduate students are required to read research articles as part of their course work. These students were all enrolled in the same course, and were paid a compensation for taking part in this project.

The task given to the participants consisted of reading the e-articles in order to self-inform, and gain sufficient knowledge so that they could lead a class discussion. This type of reading task was deemed ideal for this study as it is relatively slow and demanding, involving reflection, memorization and testing of understanding (O'Hara 1996). The task was also chosen because research showed that readers annotated more if they were expected to actively participate in the discussion rather than simply being physically present (Marshall et al. 1999; Wolfe 2001). The main research question then was:

RQ: What major *forms* of markings are recorded by e-document readers to structure their markings when working in an *individual* or *document sharing environment*, and *how do they differ* within these two environments?

Once we have identified the spectrum of marking possibilities, the following two hypotheses were tested to address the above research question:

H1: There will be a statistically significant difference between the means of the types of user markings across the various electronic documents.

H2: The mean scores of user marking forms for the group in a private environment will not differ significantly from the mean scores of the group that worked in a virtual and asynchronous document sharing environment.

2.2 Methodology

To achieve these research objectives, we employed a between-subjects design where the participants were randomly divided into two groups, group I and group S. The treatment imparted (independent variable) during this experiment was the environment, while the navigational activities were the dependent variables. In this case, the treatment (environment) had two levels (Martin 1991), reading in individual and document sharing setting, each of which was imposed on one of the two groups. This design was consistent with the post-test only, non-equivalent control group design in a quasi-experimental setting, where the control group is also regarded as one of the treatments so that both groups have one treatment each (Campbell and Stanley 1972).

For group I, nine students were recruited and each student read the allocated three articles according to the weekly class schedule, thus generating 27 articles. The *same* three articles were read by the nine participants of group S, where each participant read those three articles in a document sharing environment, consisting of multiple rounds, as per the schedule shown in Figure 1. For example, in the first round, the three articles were randomly assigned to three members of group S so that users 1, 4 & 7 got Article A to read in the first round of Week 1. The three read, and possibly marked, electronic articles obtained after the first round. The articles were then reassigned to three other participants in the second round, and this trend continued till the fourth round where the initiating readers got to review the markings made by other participants, and had a chance to add, delete, or edit the markings. This concluded the shared reading schedule for Week 1 and the project moved into Week 2 and 3, with the only changes being made in initiating reader and the article.

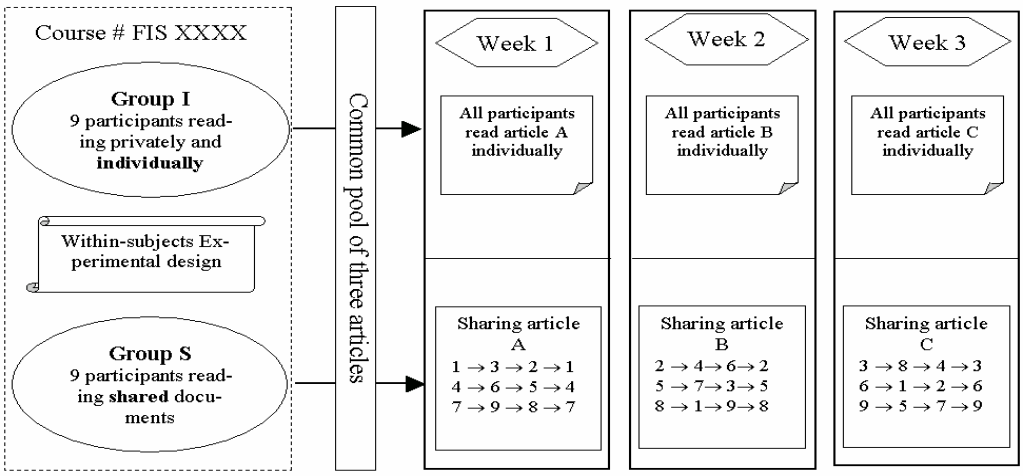

Figure 1: The research methodology

Adobe's Acrobat was the main instrument for this study and all participants, novices or experts, were required to read and mark electronic documents (in PDF-Portable Document Format) by specifically using this software. Simultaneously, user interaction with Acrobat was observed and logged by using Lotus/IBM's ScreenCam.

The first hypothesis was answered after the building of taxonomy of marking forms, from which we then identified the marking forms that differed significantly from the other types. A simple count of various types and structures of markings found in each article (Marshall 1997; Nist and Kirby 1989) and aggregated for all copies of

that document was undertaken. The resulting marking taxonomy was used in demonstrating the extension of a document's ontology to include user markings.

All the experimental arrangements were designed to maintain the user's regular reading patterns and their natural settings so as to distance the researcher and research related aspects from the actual intensive reading process. To ensure this, users undertook their readings alone and unaccompanied in a public computer lab.

3. Results and discussion

In this section we will discuss the characteristics of our participants and will list a set of observations about the users' overall marking patterns across the e-documents. We will also present the results from a series of statistical tests on the markings data that we gathered during the observation of this human-document interaction (HDI).

3.1 Profiling our participants

The questionnaire responses indicate that our participant pool comprised of 11 females and 7 male graduate students divided randomly into two groups to form the individual and shared reading environments. Most of them were in the age bracket 26-35 years (50%) and were divided almost equally between the Library/Information Sciences stream and the Information Systems stream, the two main fields of study in the Faculty of Information Studies. Almost all of participants reported daily computer usage and approximately 70% had more than five years of experience in using word processors. However, many of them had little or no experience with the somewhat advanced features of word processors. For example, approximately 70% of our participants reported that they had not used, or rarely used, the "Table of Contents" building feature.

When asked to describe the marking habits when reading paper documents, the participant responses point to underlining as the most frequently carried out activity (M=2.29). Second position was jointly secured by highlighting (M=2.53) and taking separate notes (M=2.53). This was followed by the act of making annotations in margins (M=2.88), and drawing pictures or other symbols/figures (M=3.53) on the document was the least performed activity out of these five.

3.2 Marking forms & Characteristics

During the experiment, the main factors that we focused on were the types of marking that the users made during their interactions with e-documents. From the data collected during the user studies, we were able to derive a taxonomy of the basic forms and characteristics of the user markings (anchors). This taxonomy takes into account the two environments in our experiment, i.e., when the users were reading individually (in a private environment), and when they were reading in a shared setting.

The key marking forms and characteristics observed during the analysis of the data collected during this study can be divided into three main categories:

1) Base Markings

   Base markings consisted of the most common forms of markings used on documents, for example, highlighting, underlining, and then a combination of these two in the form of compound markings.

2) Annotations

   Many diverse types of writings were termed as being a part of the annotations grouping, for example, this category included sentences that were either short in length, or lengthier ones in the form of notes that contained two or more sentence. Also discussed under this category were short annotations that took up the form of cryptic expressions, for example, "yes", "no", "true", "false", etc. Lastly, markings made to support those annotations, such as, question marks ("?"), exclamation marks ("!"), or numbers, were also included in this category.

3) Symbols and Indicators

The last of the marking anchors included all the observed symbols and indicators that did not qualify as being base markings or annotations. These included circles, straight or curved lines, arrows or arrowheads, linking or marginal bars, brackets, asterisks or stars, or pictures.

The results from the compiled data, shown in Table 1, established *underlining* (M = 8.3) as the most popular anchor used for marking text when the users were working individually, followed by *highlighting* (M = 7.1). As was seen earlier from the questionnaire data, this trend is in agreement with the user reported activities as were reportedly carried out on paper. However, these online marking trends showed a reversal in the document sharing environment where *highlighting* took over the top spot (M = 21.6) followed by *underlining* (M = 13.4). Nevertheless, despite this reversal, a clear fact that emerged from this taxonomy was that underlining and highlighting were the dominant form of markings in e-document readings, suggesting that user interface designers should focus on giving prominence to tools that facilitate these actions.

Table 1: Forms and characteristics of user markings

Marking forms & characteristics *(Anchors)*	**Individual** *(Mean/doc)*	**Doc sharing** *(Mean/doc)*
Base Markings	**(58%)**	**(88%)**
Highlighting	7.1	21.6
Underlining	8.3	13.4
Compound markings	0.4	2.2
Annotations	**(10%)**	**(7%)**
Short Sentences	0.7	1.9
Cryptic Expression (1-2 words)	1.2	0.5
Note (Three or more sentences)	0.1	0.6
Other Marks (?, !, numbers)	0.5	0.1
Symbols/Indications	**(32%)**	**(5%)**
Circle	0.9	0.4
Square	0.1	0.6
Asterisk/stars	0.4	0.1
Other pictures (Arrows, arrow-heads, drawings, lines)	0.6	0.8
Single bars in margin	5.1	0.2
Double bars in margin	1	0.0
Bars on either side of text in margin	0.5	0.0

An analysis of other e-document marking categories in individual working environment showed that making *symbols* on e-documents dominated making *annotations*. This finding is contrary to what the users reported while outlining their marking activities on paper documents, but is in line with Marshall's findings (Marshall and Brush 2004), and could perhaps indicate the lack of provision of good note making or annotating facility in electronic environment.

The annotating practice dropped further in document sharing environment and most of the slack for making symbols was taken up by base markings, indicating that people were less willing to do anything other than just place base anchors, such as, highlights and underlines while working in document sharing environments. Significant exceptions do remain as it can be observed that making of the short sentences and notes in shared environments does go up, indicating user preferences for annotating despite the lack of software support indicated by earlier results.

The overall taxonomy of these forms suggested that, other than the commonly used anchors, individual user preferences existed in terms of the types of marks that they used in their everyday readings, demonstrating the importance of understanding these markings to build more focused user profiles in client-focused systems.

To test for difference between the means of the different markings forms and strategies, two types of statistical tests were carried out in SPSS, namely:

1) One sample t-test: T-test of the type 'one sample t-test' was initially used to find significantly differing marking types. There was no treatment involved in this case, and the procedure tested whether the mean of a single variable differed from a specified constant. Thus, the two-tailed t-tests were conducted to gain knowledge about an unknown population (Gravetter and Wallnau 1999).

2) ANOVA: One way analysis of variance F-test (ANOVA) was the other approach employed to check for significant differences existing within the means of the various marking types observed in this study. The F-test checks the assumption that the samples came from populations whose means were equal. Thus, each of the marking types was compared against the others to find out if there were any significantly different marking types.

Table 2:Taxonomy of the significantly different markings forms

Marking forms & Characteristics *(Anchors)*	**Individual**			**Doc Sharing**		
	M/Doc	*t*	*Sig. (p<0.05)*	*M/Doc*	*t*	*Sig. (p<0.05)*
Base Markings						
Highlighting	7.1	1.78	0.114	**21.6**	**3.05**	**0.016**
Underlining	**8.3**	**3.48**	**0.008**	**13.4**	**2.92**	**0.019**
Compound markings	0.4	1.42	0.194	2.2	1.63	0.142
Annotations						
Short Sentences	**0.7**	**2.44**	**0.040**	**1.9**	**3.12**	**0.014**
Cryptic Expression (1-2 words)	1.2	1.29	0.232	0.5	1.90	0.094
Note (Three or more sentences)	0.1	1.41	0.195	0.6	1.81	0.108
Other Marks (?, !, numbers)	0.5	1.58	0.154	0.1	1.51	0.169
Symbols/Indications						
Circle	0.9	2.14	0.065	0.4	1.84	0.104
Square	0.1	1.00	0.347	0.6	1.58	0.154
Asterisk/stars	0.4	1.10	0.304	0.1	1.51	0.169
Other pictures (Arrows, arrow-heads, drawings, lines)	0.6	1.25	0.247	0.8	1.18	0.271
Single bars in margin	5.1	1.90	0.094	0.2	1.00	0.347
Double bars in margin	1	1.55	0.160	0.0	--	--
Bars on either side of text in margin	0.5	1.36	0.211	0.0	--	--

The first of the two tests specified above, one sample t-test, ensured that a comparison of the means for each of the marking forms was carried out against the population mean, which is unknown and hence is automatically assigned a test case value of zero by SPSS. This constant value comparison approach was chosen as it made a similarity assessment between the marking forms possible as all marking forms were being tested against the same value. Moreover, to reduce the chances of occurrences of type I error, the value of α was kept at 0.05 or at 95% confidence level and the results of these t-tests are shown in Table 2. Table 3 shows that within the markings forms, *underlining* was performed significantly more often as compared to the hypothesized mean in both the individual ($t(8) = 3.48$, $p<0.05$) and document sharing ($t(8) = 2.92$, $p<0.05$) environments, while *highlighting* differed significantly ($t(8) = 3.05$, $p<0.05$) only in the case of document sharing markings. Similarly, for annotations, *short sentences* were also marked statistically significantly more often ((P) $t(8) = 2.44$, $p<0.05$, (S) $t(8) = 1.9$, $p<0.05$) in both the environments. No other significantly different marking forms were observed.

To confirm the existence of differences between the marking types, ANOVA was carried out. This test confirmed the existence of at least one statistically different population of marking type in both individual ($F = 3.38$, $p<0.05$) and document sharing environments ($F = 7.65$, $p<0.05$). As the F-test did not tell us which of the populations differed, a set of techniques, called the post hoc comparisons were then carried out. As there is little agreement between social science researchers on which test to use for post hoc assessments (Argyrous 2000, p.364), a popular comparison, Tukey Honestly Significantly Different (HSD), was undertaken. This test revealed that in both the environments, *underlining* differed statistically significantly from almost all of the other marking types, while in document sharing environment, *highlighting* also differed statistically significantly when compared to the other marking types. The differences for the underlining are shown in Table 3.

Table 3: Underlining as a statistically significantly different marking type

Marking type #1	Marking type #2	Individual		Document Sharing	
		Mean Diff. (1-2)	Sig. (p)	Mean Diff. (1-2)	Sig. (p)
Underlining	Highlighting	1.22	1.000	-8.11	.424
	Compound marking	7.89	**.019**	11.22	**.048**
	Short Sentences	7.56	**.031**	11.59	**.034**
	Cryptic Expression (1-2 words)	7.07	.061	12.93	**.009**
	Note (Three or more sentences)	8.19	**.012**	12.85	**.010**
	Circle	7.78	**.022**	13.37	**.006**
	Square	7.37	**.041**	13.00	**.008**
	Asterisk/stars	8.15	**.013**	12.85	**.010**
	Other pictures (Arrows, arrow-heads, drawings, lines)	7.85	**.020**	13.37	**.006**
	Single bars in margin	7.74	**.024**	12.67	**.012**
	Double bars in margin	3.22	.959	13.22	**.007**
	Bars on either side of text in margin	7.26	**.047**	13.44	**.005**
	Circle	7.78	**.022**	13.44	**.005**

Based on the results tabulated in Table 2 & Table 3 for the marking forms, H1 can now be accepted and it can be said that statistically significant differences do exist between the various types of marking forms and characteristics.

Next, we will test for difference in marking trends within our two working environments, i.e., the private and shared settings, and for this we revert to our H2. To test this hypothesis, we used the Levene's test for equality of variances in SPSS, or what is commonly known as the F-test. We redraw Table 1 so as to include the results generated from our F-test. The resulting new table is now shown as Table 4 where the F-test results (not shown) indicated which t-value we should use to check for significance. Table 4 then lists the resulting t-values and their significance values.

For the marking forms, only when we drop our significance level bar down to the 90% confidence level do we see that users highlight ($t=1.78$, $p=0.1$) significantly more in a sharing environment compared to a setting when private reading takes place. Dropping our bar a little further, we see that sharing readers also tend to make shorter markings whereas, the individual readers have a bit more tendency to place side-bars in the margins. What is clear from this analysis is the fact that readers in individual and shared conditions tend to form their markings in a manner that does not vary significantly across the two settings.

Table 4: Identifying the significantly different marking forms within environments

Marking forms & characteristics *(Anchors)*	**Private** *(Mean/doc)*	**Doc sharing** *(Mean/doc)*	**t**	**Sig. (p<0.1)**
Base Markings				
Highlighting	**7.1**	**21.6**	**1.78**	**0.098**
Underlining	8.3	13.4	0.99	0.340
Compound markings	0.4	2.2	1.30	0.211
Annotations				
Short Sentences	0.7	1.9	1.67	0.115
Cryptic Expression (1-2 words)	1.2	0.5	-0.72	0.485
Note (Three or more sentences)	0.1	0.6	1.43	0.172
Other Marks (?, !, numbers)	0.5	0.1	-1.34	0.217
Symbols/Indications				
Circle	0.9	0.4	-0.97	0.346
Square	0.1	0.6	1.10	0.288
Asterisk/stars	0.4	0.1	-0.91	0.377
Other pictures (Arrows, arrow-heads, drawings, lines)	0.6	0.8	0.28	0.783
Single bars in margin	5.1	0.2	-1.81	0.107
Double bars in margin	1	0.0	-1.55	0.160
Bars on either side of text in margin	0.5	0.0	-1.36	0.211

Based on the evidence we have seen so far, we cannot reject H2, and we conclude by saying that the mean scores for the marking forms of user markings for the e-document readers reading in a private environment do not differ significantly from the mean scores of the users working in a document sharing environment. This is because we noted that only a few of the individual variables differ within the two environments. However, we did note an exception in Table 1 when we observed that as a group, symbols or indicators are used less frequently in shared environments.

At this point, the query posed by our research question is satisfied as we have identified the significant forms of markings. We have also identified the forms of markings which differ, or do not differ, statistically significantly between the users who read individually, and the readers who read e-documents in a shared environment.

3.3 Limitations

Our study had several limitations which restrict generalizability. First, we were limited to the use of one computer application in our exploration of the interaction process, as this was felt necessary to ensure that a common environment and a universal set of marking tools would be available to all readers. A second limitation was that through the introduction of the active reading process using an electronic reading/editing system, we were intervening in a user's regular reading practice. By doing this, we expected to get a better picture of how they interacted with electronic documents during the reading and marking processes. Lastly, we conducted our study on a limited number of student readers reading for a specific task. A different task in a different setting with a larger participant pool may influence the results. We realize that while the above limitations do restrict the generalizability of this study, the results do satisfactorily portray the text marking patterns of knowledge readers accessing electronic documents.

4. DEVELOPING A MARKING ONTOLOGY FOR EFFICIENT INFORMATION RETRIEVAL

This investigation was undertaken because we expect users working in a digital library to encounter an environment that allows them to assume the role of authors as they read. One of the implications that we discuss

here is the enhanced information retrieval potential through the development of an ontology. For this purpose, we use the concepts of semantic web that have been envisaged by Tim Berners-Lee (Berners-Lee and Fischetti 1999; Berners-Lee, Hendler, and Lassila 2001). Semantic Web is understood to be the instrument that will ultimately bring structure to the meaningful content of Web pages by extending the current web in such a way so that the information contained within each web page is given a well-defined meaning. This will enable computers and people to work in cooperation by sharing information, mainly because this organized structure will facilitate software agents. Thus, an environment will be created where a roaming agent will visit various web pages to carry out sophisticated information retrieval tasks assigned to it by its masters (the World Wide Web users).

Central to making the semantic web machinery work together in a seamless manner is the use of an ontology, which in context of knowledge sharing has been defined as being the specification of a conceptualization (Gruber 1993). While a taxonomy defines classes of objects and relations among them, an ontology goes much further as it is a formal, explicit specification of a conceptualization. Thus, a typical kind of ontology for the Web has a taxonomy and a set of inference rules. This is because an ontology is an abstract model of a particular field of knowledge; it describes concepts, attributes of concepts, and the relationships between concepts providing an agreed and shared vocabulary of terms and relations, as a consensual knowledge accepted by a large group of people.

From the above discussion, we can infer that once we know the user marking types, a basic ontology for information exchange utilizing these e-document markings can be developed. The marking forms and characteristics presented in Table 1 can be characterized as a taxonomy of marking activities that users are expected to carry out while reading e-documents. What remains is to convert this taxonomy into an ontology, which users of e-documents working in the web's asynchronous working environment can exploit, perhaps through agents, when they mark those documents.

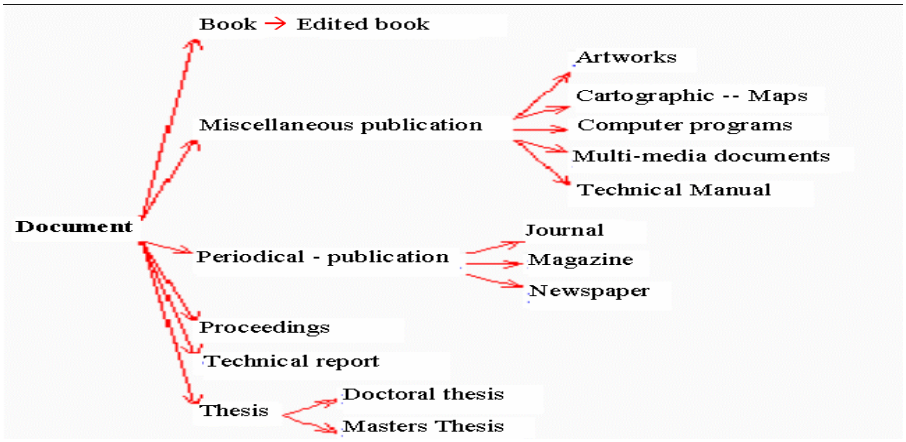

Figure 2: KSL's representation of a document through an ontology

A relevant ontology for the purposes of this research is the document ontology (*KSL Network Services* 2004), shown in Figure 2, which is part of the Ontolingua ontology library and is derived from SHOE[Φ] (Simple HTML Ontology Extensions) document model. This ontology is based on the knowledge interchange format (KIF), which is a language for describing knowledge. KIF classifies the different instances of a document by borrowing its concepts from the Structuralist Dublin Core resource types and PubMed document classifications. Examples of these instances are books, periodicals, reports, theses, etc. Each of these classes (or concepts) can be further

[Φ] SHOE is an extension to HTML that allows web page authors to place annotations on their web documents. These annotations then act as machine-readable information.

subdivided into subclasses, and the subclasses are characterized by properties and relationships, such as, author, publisher, date of publication, name of the book, journal or conference, etc

Within the range of document types shown in Figure 2, the journal type is of relevance in our research as all electronic documents used by participants were journal articles. Thus, an extension of the journal article ontology to include the marking forms is shown in Figure 3.

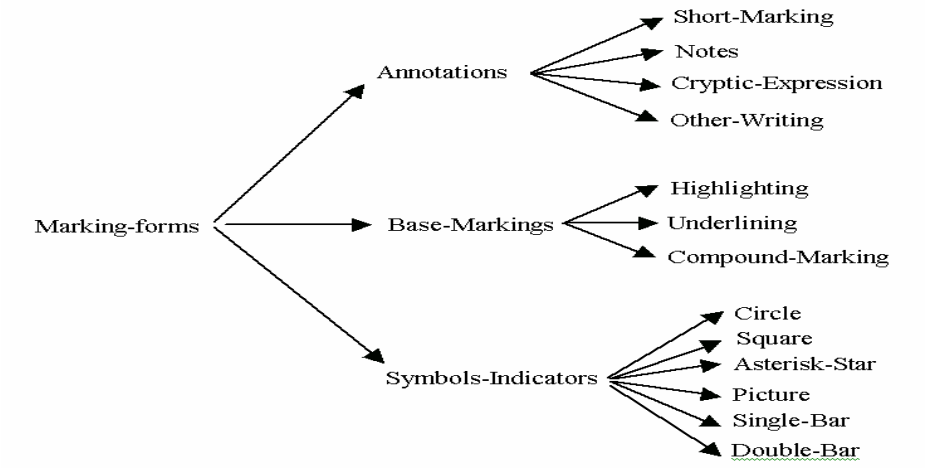

Figure 3: Marking ontology for e-documents

The root class for the marking ontology is 'Marking-Forms', which acts as a subclass to all subclasses of the document type 'journal'. There are three main subclasses to 'Marking-Forms' ('Annotations', 'Base-Markings', and 'Symbols-Indicators'), and each of these subclasses may be drilled down further to form a bottom conceptual layer of subclasses. Thus, while Base-markings are middle level concepts, they can be divided into three subclasses comprising the bottom layer; highlighting, underline, and compound marking. Each class in the bottom layer will have some properties or slots, for example, highlighting may have different colors, thickness, or length. The slot-type, range, and slot-cardinality will then limit the values these slots can own. As a result of this classification, each marking placed on an electronic journal article by its user can be automatically characterized as being an instance of the marking forms ontology, will be owned by a class, and will have appropriate slots. Thus, every marking placed on an e-document can now be automatically characterized by this ontology, and recognized as such by a free roaming web agent.

The sets of markings shown in Figure 3 are essentially ontological concepts, which can be defined by Resource Discovery Framework (RDF)-based metadata for actual implementation. Each piece of marked information will then potentially lead to the creation of a new entry in an ontologically represented data store, and RDF can be used to specify the classes in the ontology, similar to techniques used by ontology based knowledge sharing system such as OntoShare (Davies, Duke, and Sure 2003). RDF can also be used to populate the ontology with instances, as more information within articles is marked and made available for sharing. Thus, the actual knowledge is represented in an RDF or XML-based ontology, which caters for the markings made on electronic documents according to the taxonomies that were developed earlier, and uses an implementation structure similar to W3C's Anntoea (Kahan et al. 2001).

The proposed ontology-based framework should greatly extend the electronic journal article's reorganization and recognition because the source text, marked-up by the readers of that document, would also be available for searching by search engines, or by web agents. This availability of the marked text in-turn will improve the

document search and retrieval results, because the marked up content of the source text would also be available within the semantic web search sphere. Thus, the current searching trends of looking for keywords within the title, abstract or indexing can be extended. Masters (users) of the search agents could then specify a search criterion for looking within the marked content, or within annotations on electronic documents, in addition to searching the source text.

5. CONCLUDING THOUGHTS

This paper presented a taxonomy of the possible marking types that users, working with electronic documents, can make. We discussed these forms and characteristics of markings with a focus on the users working in private or shared environments. Also demonstrated was the conceptual framework of how the developed taxonomy can be used to extend an existing ontology of documents, which should result in efficient information retrieval. We argue that, our research goes towards helping in narrowing the gap between the e-document readers and software developers, as we believe that an understanding of the user's action at the system level is the best way to adapt a user-centric system development approach. We intend to investigate further the role played by the users in the process of markings and knowledge sharing, as well as will look closer at the functions performed by the user markings and how they are perceived by the readers.

6. ACKNOWLEDGMENTS

This research was made possible by a research grant from International Council for Canadian Studies.

7. REFERENCES

Adler, A., A. Gujar, B. L. Harrison, K. O'Hara, and A. Sellen. 1998. A diary study of work-related reading: design implications for digital reading devices. Paper read at Proceedings of the Conference on Human Factors in Computing Systems, Apr. 18 - 23, at Los Angeles, CA.

Argyrous, George. 2000. *Statistics for social and health research: With a guide to SPSS*. Thousand Oaks: Sage publications.

Berners-Lee, T., and M. Fischetti. 1999. *Weaving the Web : the original design and ultimate destiny of the World Wide Web by its inventor*. 1st ed. San Francisco: HarperSanFrancisco.

Berners-Lee, T., J. Hendler, and O. Lassila. 2003. *The Semantic Web* [WWW]. ScientificAmerican.com 2001 [cited November 10 2003].

Brush, A. J. B., D. Bargeron, A. Gupta, and J. J. Cadiz. 2001. Robust annotation positioning in digital documents. Paper read at Proceedings of the SIGCHI '01, at Seattle, WA.

Campbell, D. T., and J. C. Stanley. 1972. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally & Co.

Coleridge, Samuel Taylor. 1985. *Marginalia*. Edited by H. J. Jackson. 12 vols. Vol. 12, pt. 1-6. Princeton: Princeton University Press.

Davies, John, Alistair Duke, and York Sure. 2003. OntoShare – A Knowledge Management Environment
for Virtual Communities of Practice. Paper read at Proceedings of the ACM's 2nd International Conference on Knowledge Capture, October 23–25, at Sanibel Island, Florida, USA.

Fowler, R. L., and A. S. Barker. 1974. Effectiveness of highlighting for retention of text material. *Journal of Applied Psychology* 59 (3):358-364.

Gravetter, Frederick J., and Larry B. Wallnau. 1999. *Essentials of statistics for the behavioral sciences*. 3rd ed. Pacific Grove: Brooks/Cole Publishing Co.

Gruber, T. R. 1993. A translation approach to portable ontologies. *Knowledge Acquisition* 5 (2):199-220.

Idstein, P., and J. R. Jenkins. 1972. Underlining versus repetitive reading. *Journal of Educational Research* 65:321-323.

Jackson, H. J. 2001. *Marginalia: Readers writing in books*. New Haven & London: Yale University Press.

Kahan, J., M-R. Koivunen, E. P. 'Hommeaux, and R. R. Swick. 2001. Annotea: An Open RDF Infrastructure for Shared Web Annotations. Paper read at Proceedings of the 10th International Conference on World Wide Web, WWW10, May, at Hong Kong.

KSL Network Services 2003. [World Wide Web]. Stanford University, Knowledge Systems Laboratory 2004 [cited Nov 10 2003]. Available from http://www-ksl-svc.stanford.edu:5915/.

Marshall, C. C. 1997. Annotation: from paper books to digital library. Paper read at Proceedings of the 2nd International Conference on Digital Libraries, July 23-36, at Philadelphia, PA.

Marshall, C. C., and A. J. B. Brush. 2002. From Personal to Shared Annotations. Paper read at Proceedings of the CHI 2002 Conference on Human Factors in Computing Systems., April 20-25, at Minneapolis, MN.

———. 2004. Exploring the relationship between personal and public annotations. Paper read at IEEE/ACM Joint Conference on Digital Libraries, June 7-11, at Tucson, AZ.

Marshall, C. C., M. N. Price, G. Golovchinsky, and B. N. Schilit. 1999. Introducing a digital library reading appliance into a reading group. Paper read at Proceedings of the Fourth ACM International Conference on Digital Libraries, Aug. 11-14, at Berkeley, CA.

Martin, D. W. 1991. *Doing psychology experiments*. 3rd ed. Pacific Grove, CA.: Brooks/Cole Publishing Co.

Neuwirth, C. M., D. S. Kaufer, R. Chandhok, and J. H. Morris. 1990. Issues in the design of computer support for co-authoring and commenting. Paper read at Proceedings of the Conference on Computer-supported Cooperative Work, Oct. 7-10, at Los Angeles, CA.

Nist, S. L., and K. Kirby. 1989. The text marking patterns of college students. *Reading Psychology* 10:321-338.

O'Hara, K. 1996. Towards a Typology of Reading Goals. Cambridge, UK: Rank Xerox Research Centre, Cambridge Laboratory.

Stoddard, R. E. 1985. *Marks in books, illustrated and explained*. Cambridge, MA: Harvard University Press.

Tribble, E. B. 1997. "Like a looking-glas in the frame": From the marginal note to the footnote. In *The margins of the text*, edited by D. C. Greetham. Ann Arbor, MI: University of Michigan Press.

Wenger, E. 1998. *Communities of Practice: Learning, Meaning and Identity*. New York: Cambridge University Press.

Wolfe, J. L. 2001. Pedagogical uses of annotations and annotation technologies. PHD Dissertation, University of Texas, Austin.

Wolfe, J. L., and C. M. Neuwirth. 2001. From the margins to the center: The future of annotations. *Journal of Business and Technical Communication* 15 (3):333-371.