

Reader's Preferences in the Formats of Web-based Academic Articles

Y. Rho

School of Computer Sci. & Eng.
University of New South Wales
yrho@cse.unsw.edu.au

T.D. Gedeon

School of Information Tech.
Murdoch University
tdgedeon@murdoch.edu.au

C.K. Kim

Department of Computer Sci.
University of Incheon, Korea
ckkim@lion.incheon.ac.kr

Abstract

No standard format exists for the many academic articles available on the Web and little is known about user reading patterns. This paper explores these issues using data from two online surveys: one email-based, the other Web-based. Our results suggests that people take an overview from the screen, and then, if they are interested in an article, print it out in order to read it properly. The simple two-frame format was regarded as the best by 47% of the respondents, whereas the cascade page format was regarded as the worst by 65% of the respondents. Interestingly, 26% considered the paper-like format, widely used in Web-based articles, to be the worst. Different results were obtained when interactive examples were embedded in the survey.

Keywords: Web-based articles; Reading patterns and formats; Web-based survey; Digital library

1. Introduction

A number of Web sites exist that publish academic articles. There are no well-developed formatting guidelines for these sites and so a variety of different formats are in use. This paper presents survey information which should be useful in determining how best to format these sites.

Articles on the Web

One format being widely used on the Web is shown in Figure 1. This format is very similar to its paper-based counterpart. We chose some well-known sites in the HCI area: conference sites such as WWW6, WWW8 and ACM CHI97, and journal and magazine sites such as ACM SIGCHI Bulletin, International Journal of Human-Computer Studies (IJHCS), Alertbox and ACM Transactions on Computer Human Interaction (TOCHI). shows some features of article formats presented on those sites. These features can be summarised as follows:

- Abstracts are always in the articles; tables of contents (TOC) are sometimes included;
- Scrolling is a common method of navigation;
- If indexing based on a TOC exists, sliding is used as a secondary method at some sites;
- The single window layout is most popular; and
- The ACM TOCHI seems to use the Web only as a delivery medium.

The formats of those sites are different as described. This may be due to the **designers'** different assumptions on user reading patterns. Two online surveys were carried out to investigate reading patterns and formats. Our survey participants were volunteer researchers in information technology and related areas.

Table 1 : Various formats of legacy Web-based articles

	Information overviews	Manipulation methods	Windows layout	Others
WWW6[17]	Abstract only	Scrolling	Single window	
WWW8[18]	<ul style="list-style-type: none"> • Abstract only or • Abstract & TOC or • TOC & Abstract 	<ul style="list-style-type: none"> • Scrolling only or • Sliding (Prev-Next) or • Indexing + Scrolling 	Single window	Major format is paper-like.
CHI97[10]	Abstract - TOC	Indexing + Scrolling	Single window	
ACM SIGCHI Bulletin [11]	TOC - Abstract	<ul style="list-style-type: none"> • Scrolling 	Single window	
IJHCS [14]	<ul style="list-style-type: none"> • Abstract and TOC in frames • Abstract only in single window 	<ul style="list-style-type: none"> • Downloading to print • (Indexing + sliding) frames • Scrolling 	<ul style="list-style-type: none"> • Single window • Multiple frames 	<ul style="list-style-type: none"> • Multiple formats available • Comments
Alertbox [13]	Abstract only	Scrolling only	Single window	Path on the top
ACM TOCHI [12]	Abstract and general terms	Downloading to print	Single window	Delivery purpose

Proceedings of the Fourth Australian Computing Symposium, Coffs Harbour, Australia, December 3, 1999.

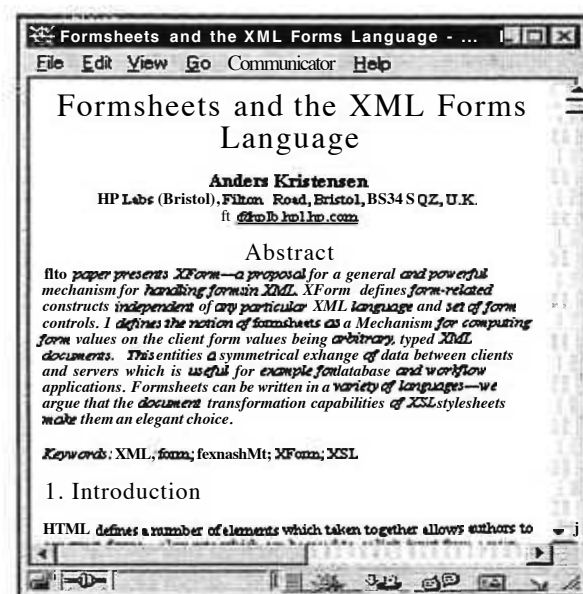


Figure 1: A popular format [15]

Online survey

The online survey (see [7][5]) is considered to be at least as good as the paper and pencil survey. In fact, it is better because it encourages participants to give more comments [9]. In the case of Web-based applications the benefit of being online is that interactive examples can be embedded in a questionnaire. These examples can provide participants with a more immediate experience of the targets in question. The use of this method is not reported as yet.

2. Email-based survey on reading patterns: the first survey

Purposes of the survey

The purpose of this survey was to determine whether researchers find research articles from the Web. If they do, what are their usage patterns?

Method

An Email-based questionnaire of four questions was distributed to 130 research people in the School of Computer Science and Engineering of the University of New South Wales in Australia. No examples were included in the questionnaire (Figure 2).

Results

We received 23 replies; that is, 18% of those 130 polled. 22 (96%) indicated that they find articles from the Web.

Q1: Do you find academic articles from the Web? (Yes / No)

If 'Yes' for Q1,

Q2: What describes your behaviour best when you have an article on the Web?

- (1) You just print it out, and then read the printed article.
- (2) You read the first few lines on the first screen, print out the article if you are interested in it, and read the printed article.
- (3) You read some concise parts such as titles & abstracts, print out the article if you are interested in it, and read the printed article.
- (4) You scan through the article, print it out if you are interested in it, and then read the printed article.
- (5) You read the article from the screen.
- (6) Others (please describe)

Q3: What could be your second choice in Q2?

If 'No' for Q1,

Q4: Why?

- (1) The Web articles are NOT credible.
- (2) The Web articles are frequently updated.
- (3) Others (please describe)

Figure 2: Questions in the email questionnaire

Table 2: usage patterns with Web articles

	Details	1 st (Q2)		2 nd (Q3)	
		Frequency	Frequency	Frequency	Frequency
Pattern 1	Print and read	1	5%	1	5%
Pattern 2	Read from the first screen, print and then read	0	0%	6	30%
Pattern 3	Read concise parts, print and then read	14	64%	6	30%
Pattern 4	Scan through, print and then read	7	32%	4	20%
Pattern 5	Read from the screen	0	0%	3	15%
Others		0	0%	0	0%
N		22	100%	20	100%

Table 2 shows the responses regarding usage patterns. For Q2 in Figure 2, 64% had Pattern 3 as their first choice and 32% had Pattern 4 as their first choice. The other responses were not significant. An interesting result is that no one selected Pattern 2 as their first choice. For the second choice, Patterns 2 and 3 together recorded the highest selection frequency with six (30%).

Discussion

The most common usage pattern for Web-based articles is Pattern 3. However, Pattern 4 should not be ignored as it is the first choice for 32%. These responses are very similar to the results regarding usage patterns of paper-based academic journals shown in [1]. This may be due to the fact that the reader's reading behaviour is guided by the metastructure they are used to (see [2] [3] for metastructure). This metastructure seems to guide

article readers to the reading process shown in Figure 3. The respondents who prefer Pattern 3 appear to want to view the concise parts on the first screen. On the other hand, the respondents who prefer Pattern 4 seem to focus much more on the overall contents of an article than on its interface.

Survey conclusion

Those readers who find articles on the Web mostly take an overview from the screen, after which they print them out and read them. They seldom read entire articles from the screen. So which formats then do readers prefer and which do they dislike? The second survey answers this question.

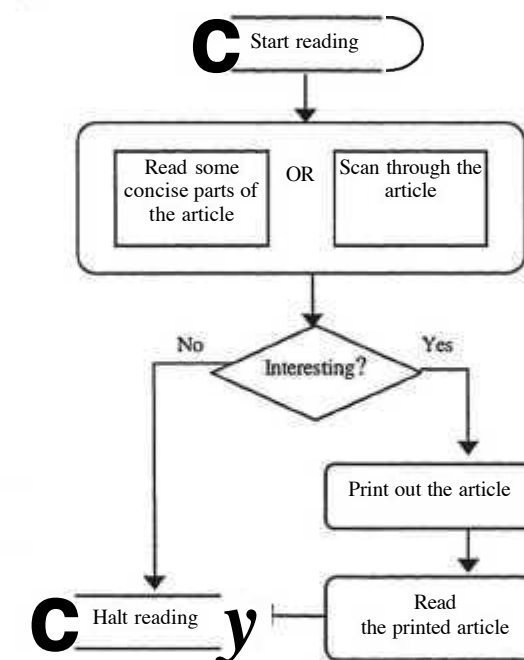


Figure 3: A reading pattern from the 1st survey

3. Web-based survey on reading patterns and interface formats: the second survey

The previous survey revealed the common reading pattern. A Web-based survey, based on this knowledge, was also conducted. Examples were provided to the participants in a Web-based questionnaire, consisting of checklist style questions. Figure 6 shows a partial screen shot of the Web-based questionnaire used in the survey.

Purposes of the survey

The second survey had two purposes. One was to identify which format readers prefer and how they use it. The other was to investigate the effects of interactive examples in the questionnaire.

Methods

A Web-based questionnaire consisting of 18 questions (i.e., [6]) was developed. The first part of this questionnaire concerns the environment settings. This part controls the visual properties and the window size of a browser in terms of the amount of information amount. This helps to avoid visual volume effects when presenting interactive examples.

The second part consists of questions about three different layers: overview types, window layouts and manipulation methods. The data from this part is not discussed in this paper. The last part is about usage patterns and overall preferences in interface formats. Each question has at least one corresponding example link.

The URL for the survey questionnaire was sent by email to 150 researchers in information technology and related areas. They were research students, research staff and academic staff. Undergraduate and coursework students were not included because they seldom use the Web to find academic articles.

Results

We received 34 replies; i.e. 23% replied to the questionnaire. Most of them used 17" monitors to complete the questionnaire (Mean=17.1").

Table 3: Reading patterns

	Details	1 st choice		2 nd choice	
		Frequency	Frequency	Frequency	Frequency
Pattern 1	Print and read	2	6%	2	6%
Pattern 2	Read from the first screen, print and then read	6	18%	10	29%
Pattern 3	Read concise parts, print and then read	19	56%	8	24%
Pattern 4	Scan through, print and then read	6	18%	10	29%
Pattern 5	Read from the screen	1	3%	4	12%
Others		0	0%	0	0%
N		34	100%	34	100%

Like the previous email-based survey, Table 3 shows that Pattern 3 (at 56%) is the most common of the reading patterns. Patterns 2 and 4 (both 18%) were the second most common. In the previous survey, there was 0% for Pattern 2 and 32% for Pattern 4. Obviously, there is a large discrepancy between the results of the two surveys. The possible reason for this difference will be discussed in the Discussion section

Table 4: Preferences in overall formats

	Best		2nd		3rd		4th		Worst	
	Frequency	Frequency	Frequency	Frequency	Frequency	Frequency	Frequency	Frequency	Frequency	Frequency
1 Paper-like	2	6%	7	21%	9	26	6	18	9	26%
2 Paper-like with TOC	12	35%	13	38%	5	15	4	12	0	0%
3 Two frames	16	47%	5	15%	7	21	5	15	0	0%
4 Slides	2	6%	8	24%	10	29	12	35	1	3%
5 Cascades	2	6%	1	3%	2	6	5	15	22	65%
N/A	0	0%	0	0%	1	3	2	6	2	6%
N	34		34		34		34		34	

Table 5: Correlation between reading patterns and formats for the two first choices

	Format 1			Format 2			Format 3			Format 4			Format 5			N(pattern)	
Pattern 1	0			2	<u>100%</u>	17%	0	<u>0%</u>	0%	0			0			2	6%
Pattern 2	1	<u>17%</u>	50%	1	<u>17%</u>	8%	3	<u>50%</u>	19%	0			1	<u>17%</u>	50%	6	<u>18%</u>
Pattern 3	1	<u>5%</u>	50%	6	<u>32%</u>	50%	9	<u>47%</u>	<u>56%</u>	2	<u>11%</u>	100%	1	<u>5%</u>	50%	19	56%
Pattern 4	0			2	<u>33%</u>	17%	4	<u>67%</u>	25%	0			0			6	18%
Pattern 5	0			1	<u>100%</u>	8%	0		0%	0			0			1	3%
N(format)	2	6%			35%		16	47%		2	6%		2	6%		34	100%

Table 4 shows that 47% of the respondents chose the two-frame format out of the five example formats (try Q5 in [6] for the examples). 35% selected the paper-like-with-TOC format as the best. 65% selected the cascades as the worst. It is interesting that 26% of the respondents selected the paper-like format as the worst. It should be noted that no one indicated either the paper-like-with-TOC format or the two-frame format as the worst.

Table 5 shows a correlation between the patterns and the formats. The underscored percentages of the shaded row show the format distribution over Pattern 3. Format 3's contribution to Pattern 3 is the highest at 47% and Format 2 is next at 32%. In the other view, the shaded column shown corresponds to the pattern distribution over Format 3. Pattern 3 is at 56%, followed by Patterns 4 and 2. Pattern 3 matches Format 3 and vice versa. In addition, Patterns 2 and 4 match Format 3 as well, at 50% and 67% respectively.

Discussion

The most common reading pattern for Web-based articles is for readers to look at an article in brief on the Web, and then they print it out (if they are interested in it) and then read it. The second survey produced a similar result to the first survey.

Patterns 2 and 4 revealed a big difference between the results of the two surveys. In the first survey, Pattern 4 polled 32% and Pattern 2 polled 0%. Then, in the second survey, Pattern 4 polled only 18% (a 14% decrease) while Pattern 2 polled 18% (an 18% increase). So the portions are even. Why? There was no significant difference in the distribution list and the question for this topic.

The difference seems to relate to the presence of formatting examples. The first questionnaire was based on email with no examples to try. However, the second questionnaire was based on the Web with many examples. The participants had many chances to try different formats of Web-based articles before answering the questions. They were likely to be aware of the importance of seeing the first screen. There were also some comments on that point from some participants.

The two-frame format in Figure 4 was most preferred by the respondents, which goes against the popular idea that using frames is bad. With respect to the major reading patterns the common bridge from screen to paper is printing. Therefore, article interfaces for readers have to be able to support two different media: the Web and paper. The majority (47%) chose the two-frame format as the best for the purpose and the paper-like-with-TOC format as the next (35%). 65% chose the cascade format as the worst (65%), but no one selected either the two-frame format or the paper-like-with-TOC format as the worst. The patterns and formats showed a correlation: Pattern 2, 3 and 4 most closely match the two-frame format (Table 5).

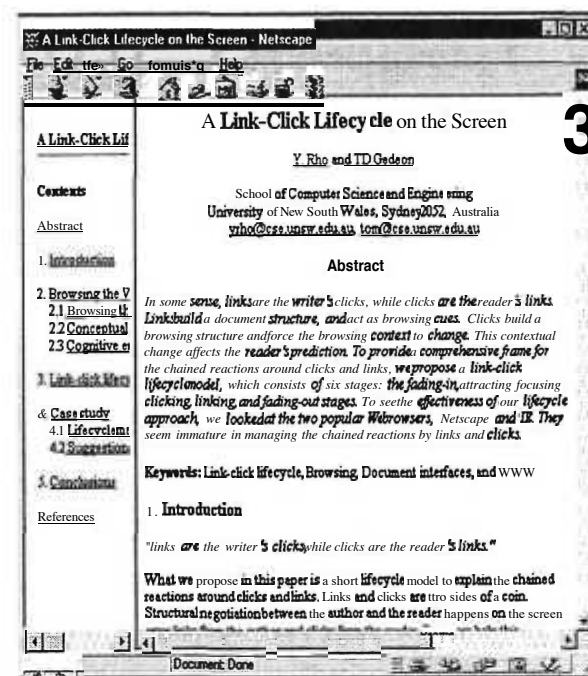


Figure 4: The two-frame format examined in the survey [16]

Survey conclusion

The early part of the user reading process, which is to take an overview of an article, happens on-screen (Figure 5). The survey results show that the early part consisting of taking an overview and printing are most

likely to be supported by the two-frame format (Figure 4), which ensures the quality of the paper-based article format when printed.

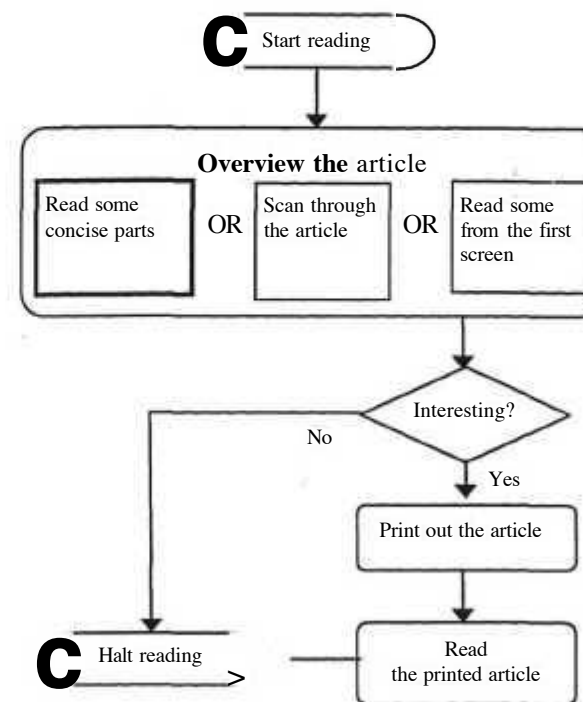


Figure 5: A typical reading process pattern with Web-based academic articles

4. Conclusion

User's reading patterns and preferences in the formats for Web-based academic articles were investigated in two online surveys: the first was email-based, the second Web-based.

The survey results show the following:

- The two-frame format (Figure 4), which consists of a TOC frame for navigation and a content frame for information, was the most favoured.
- Providing two versions together, one in the two-frame format and the other in the paper-like-with-TOC format, seems reasonable as the preference percentage (47%) for the two-frame format is not clearly dominant over the 35% of the other.
- The cascaded page format was considered by 65% to be the worst.
- An interesting discovery is that a quarter of the respondents liked least the paper-like format which is widely used.
- The use of interactive examples in the Web-based questionnaire seems to have made participants recognise the importance of the first screen.

This research focused on structural properties of Web-based academic articles from the viewpoints of information, interfaces and interactions (in short, III).

Many other features including reference to other articles, multimedia presentation and annotation need to evolve to make academic articles more readable on the Web. Further research on details should be done with real reading tasks. In addition, Nielsen's claim on the inverted pyramids ([4]) needs to be tested from the III viewpoint. Do we really have to write academic articles in the structure of inverted pyramids or can we provide navigation aids instead?

References

- [1] Dillon, A., "New Technology and the Reading Process," Computers in Libraries, July 1991, 23-26.
- [2] Dillon, A., "Readers' models of text structures: The case of academic articles," International Journal of Man-Machine Studies, 35, 1991, 913-925.
- [3] Dillon, A., "Expertise and the Perception of Shape in Information," Journal of the American Society for Information Science, 47(10), 1996, 786-788.
- [4] Nielson, J. "Inverted Pyramids in Cyberspace," Alertbox, 1996. <http://www.useit.com/alertbox/9606.html>
- [5] Perlman G., "Web-Based User Interface Evaluation with Questionnaires," 1997. <http://www.acm.org/~perlman/question.html>
- [6] Rho, Y., Questionnaire on Reader's preferences in Web-based Academic Articles, 1999. <http://www.cse.unsw.edu.au/~yrho/WebPaper/Question.htm>
- [7] Root, R. W., Draper, S., "Questionnaires as a software evaluation tool." CHI'83 Proceedings, December, 1983, 83-87.
- [8] Shneiderman, B., "Designing Information-Abundant Websites: Issues and Recommendations," International J of Human-Computer Studies, 1997. <http://www.cs.umd.edu/projects/hcil/members/shneiderman/ijhcs/main.html>
- [9] Slaughter, L., Harper, B., and Norman, K., "Assessing The Equivalence Of The Paper And On-line Formats Of The Quis 5.5". Proceedings of the 2nd Annual Mid-Atlantic Human Factors Conference, Washington, D.C, 1994, 87-91. <http://lap.umd.edu/LAPFolder/Papers/SHN.html>
- [10] ACM CHI97, 1997. <http://www.acm.org/sigchi/chi97/proceedings/paper/plp.htm>
- [11] ACM/SIGCHI Bulletin, 1997. <http://www.acm.org/sigchi/bulletin/1997.4/ross.html>
- [12] ACM Transactions on Computer-Human Interaction.

- <http://www.acm.org/pubs/contents/journals/tochi/1998-5/>
- [13] Alertbox, 1997. <http://www.useit.com/papers/webwriting/writing.html>
- [14] International Journal of Human-Computer Studies, 1997. <http://ijhcs.open.ac.uk>
- [15] Kristensen, A., "Formsheets and the XML Forms Language," WWW8 Proceedings, 1999. <http://www8.org/w8-papers/1c-xml/formsheets/formsheets.html>
- [16] Rho, Y., Gedeon, TD, "A Link-Click Lifecycle on the screen," APWeb'98 Proceedings, Beijing, 1998. http://www.cse.unsw.edu.au/~yrho/Publications/apweb98/apweb98_C-tiled_main.htm
- [17] WWW6 Conference, 1996. <http://www.scope.gmd.de/info/www6/technical/paper003/paper3.html>
- [18] WWW8 Conference, 1999. <http://www8.org/fullpaper.html>

Questionnaire on Web article usage patterns - Netscape

File Edit View Go frmmuncator Help

III. Usage patterns and overall preferences

Reading activity patterns with Web articles

Q4 What describes your behavior best when you have an article on the Web?

	Best	2ndBest
(1) You just print it out, and then read the printed article.	C	C
(2) You read some of the first screen, print out the article if you are interested in it, and then read the printed article.	C	C
(3) You read some concise parts such as titles & abstracts, print out the article if interested, and then read the printed article.	C	C
(4) You scan through the article, print it out if interested, and then read the printed article.	C	C
(5) You read the article from the screen.	C	r
(6) Others (please describe)	CD	CD

Overall comment

Preferences in overall Web article formats

Q5 Which style will support your behavior best? (please don't count speed, examine examples and order them)

	Examples	Best	2nd	3rd	4th	Worst
(1) Paper-like style of a long page	Paper-like	C	r	C	r	C
(2) Paper-like style of a long page with TOC links	Paper-like with TOC	C	r	C	C	C
(3) Two frames for TOC links and contents	Frames with TOC	C	r	C	C	C
(4) The style of slides presentation with TOC links	Slides with TOC	C	r	C	C	C
(5) Cascaded multiple pages	Cascades with TOC	C	C	C	C	C

Overall comment D

Your last comment or suggestions D

Document Done

Figure 6: A partial screen shot of the Web-based questionnaire [6]

DYNAMIC HYPER-LINKING BY QUERYING FOR A FCA-BASED QUERY SYSTEM

Bernd Groh

School of Information Technology
Griffith University
PMB 50 Gold Coast Mail Centre
QLD 9726, Australia

b.groh@gu.edu.au

Peter Eklund

School of Information Technology
Griffith University
PMB 50 Gold Coast Mail Centre
QLD 9726, Australia

p.eklund@gu.edu.au

Abstract

This paper presents a mechanism for hyper-linking documents by search-terms. Search-terms are selected by the user interactively building a formal concept lattice. In order to explain this interface we give some background to Formal Concept Analysis and an example is developed which illustrates the use of the concept lattice. Selected search-terms are used to create hyper-links, based on term repetition.

As the search-terms differ between queries, we need a mechanism from which to dynamically create the target hyper-linked HTML documents. Therefore, documents are stored in a structure which is based on a word-list rather than plain text format. The documents are represented as links between the individual words within the word-list. In so doing the word-list becomes a full-text-retrieval index into each word in each of those documents and therefore provides a good basis for the fast creation of an HTML document set from specific queries by keywords.

To have the words in a word-list from which the documents are created also allows easy classification of words which should be hyper-linked within specific HTML documents. Furthermore, both documents and hyper-linking keywords are stored as well in this structure since any word in any document is indexed by the word-list.

Keywords: Document Databases, WWW and Internet.

1 Introduction

There have been several developments in automatically generating hyper-linked documents, from hyper-linking by term repetition to semantic approaches based on similarity measures between

Proceedings of the Fourth Australasian Document Computing Symposium, Coffs Harbour, Australia, December 3, 1999.

documents. One approach by Green [7, 8] uses lexical chains [9] to measure the similarity between documents. This mechanism is based on WordNet [5] and generates groups of related words within documents. Those groups have specific meanings and can be related to groups with equivalent or similar meanings in other documents.

Our approach is query-driven and focuses on keywords. We use concept lattices as a visual interface so that the user can navigate to a specific subset of documents with given combinations of search-terms (or keywords) of interest. The hyper-links will be created based on repetition of those terms and will exist only between documents in the specific subset that satisfies the query. The technique to create the concept lattices is called Formal Concept Analysis (FCA) [6].

We structure the paper with a brief introduction to FCA. Next, based on an example of a medical document-set, we demonstrate how to read a concept lattice and how it helps the user to identify documents of interest. As document-sets differ, depending on the terms appearing in the concept lattice and on the selected concepts within the lattice, hyper-links have to be created dynamically.

In Section 4, we present a mechanism to dynamically generate a hyper-linked document-set, wrt to a specific query, based on term repetition. The approach uses a memory structure in which the entire document collection is stored and an algorithm to create the hyper-linked document-subset. The structure is based on a word-list, containing entire words and links between the words. These links, when followed, constitute the document-set.

As our interest lies in complete words which precisely match a pre-defined list of terms, in our example medical terms, fast string searching algorithms based on substrings, such as Suffix Trees, cannot be used. Our approach is explained on a small example in Section 5 and the algorithm is described in Section 6. The last sections conclude the paper and provide an outline of current and future work.