Visualising Document Content with Metadata to Facilitate Goal-directed Search

Mischa Weiss-Lijn, Janet T. McDonnell
Department of Computer Science
University of College London, London
WC1E 6BT, UK
Email - {m.weiss-lijn, j.mcdonnell}@cs.ucl.ac.uk

Leslie James
Strategy Research Team,
J Sainsbury plc, London

Much corporate knowledge is captured in documents. This paper will describe the design and evaluation of an application, named Gridvis, which uses metadata to enable rapid goal-directed searches within *corporate documents. Gridvis visualises metadata that describes and differentiates the contents of individual paragraphs. This visualisation is used as the basis for rapid querying and navigation. An experimental evaluation of Gridvis performance on representative work tasks is described. Qualitative analysis of the data suggests that there is substantial room for performance improvements, and has inspired design changes to realise this potential. Quantitative analysis proved inadequate to detect a numerical advantage due to the sample size restrictions inherent when doing academic research in a corporate setting. The problems of empirical evaluation in a corporate setting are discussed. An alternative approach to reliable evaluation is proposed.*

1 Introduction

Part of the semantic web's mission is to provide access to the nuggets of relevant information within online documents. It therefore addresses the need to make better use of knowledge in stored corporate documentation.

This paper is a part of ongoing research looking at whether goal-directed search of a document could be facilitated using an interactive visualisation of metadata that describes the content of paragraphs within corporate documents. Gridvis, the system built to test out this idea, is first described, then the results of an empirical evaluation are discussed, and finally suggestions are made for further evaluation.

The work was motivated by the need within Sainsbury's, a large UK retail organisation which sponsors this research, to make better use of the knowledge they store in corporate documents. There is much work currently looking at how to locate and automatically distribute relevant documents to employees. This research is therefore exploring a technology for improving the effectiveness of using documents once they have already been located.

Gridvis is intended to support an important class of document use where information needs are realised at the paragraph level. Adler et al [1] identify this prevalent class of reading in a study of work related document use in a range of work settings. They characterise it as 'goal-directed reading', which they described as "sampling information in the text which satisfies the goal of the search" [1]. Van House [8] concludes that "Workplace users [...] want to retrieve information rather than documents per se.". It is therefore likely that search goals in digital library use are most commonly satisfied at the paragraph or section level. Gridvis has therefore been conceived to facilitate goal-directed search by helping the user home in on paragraphs relevant to their goal.

2 System Description

Gridvis was envisaged for use with a small but important portion of a diverse corporate document landscape, which can be characterised as being information rich and having a relatively long life. A set of documents with these characteristics was selected (as described in the evaluation section) to serve as a basis for developing metadata taxonomy, and conducting evaluation.

Gridvis parses a document marked up with metadata and one containing a metadata taxonomy. It uses the result to produce a visualisation and customised HTML views of the document text.

The metadata taxonomy is necessary for several reasons, only some of which can be fully described here. For the users of Gridvis undertaking a goal-directed

0-7695-1195-3/01 $10.00 © 2001 IEEE

search, it is essential that the metadata has some meaningful organisation. With a good taxonomy the user can rapidly locate a small subset of the metadata relevant to their information need, rather then undertaking an exhaustive linear search through all the metadata tags describing the documents contents. Other benefits of the metadata taxonomy are the advantages it has for metadata authoring (discussed below), and the extra power it allows automatic searches.

The metadata taxonomy developed for Gridvis, is essentially a simplified form of the ontologies use in the Enterprise [7] and $(KA)^2$ projects [2]. These ontologies were engineered specifically to take advantage of advanced automated searching techniques for large document collections. Gridvis on the other hand is for the manual search of single documents. So, although it would in principle be able to use metadata produced within these sophisticated ontologies, Gridvis only requires a simple taxonomy.

The metadata and the metadata taxonomy were constructed by adapting informal ontology construction methodologies [6]. The metadata was manually produced for each paragraph of the documents prepared for Gridvis, using the taxonomy as a guide, to encourage complete and consistent tagging. Each instance of metadata was given an 'applicability' attribute which recorded the degree to which the metadata tag applied to a paragraph (1 – 'is principally about', 2 'partly about' or, 3 'mentions in passing').

Since Gridvis is designed for use on a corporate intranet in a corporation where software installation for low

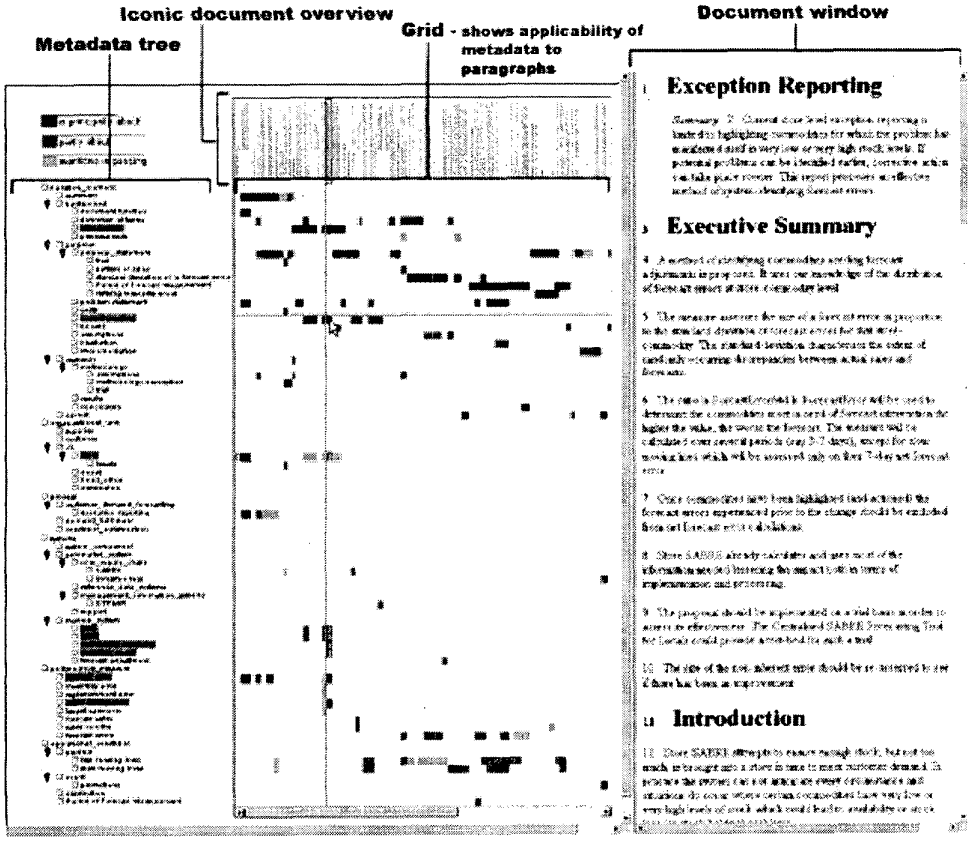

Figure 1 A screenshot of Gridvis. By placing their mouse over a cell in the grid the user has highlighted a paragraph in the iconic document overview and all the metadata terms in the metadata tree relevant to this paragraph.

priority applications is problematic it was given a client-server architecture: The visualisation is a client-side Java applet, the queries are answered by a Java servlet using XLS-T to produce customised HTML documents. The documents and taxonomy are encoded as XML. The document structure is described with a standard hierarchical XML structure, the metadata is embedded within this alongside the text it describes using RDF.

2.1 The visualization

The Gridvis client-side application contains a visualisation of a document's metadata (see fig 1). This visualisation can be considered as 3 interlinked sections; the metadata tree on the left-hand side, the iconic document overview running along the top, and the grid sitting at the centre. The metadata tree consists of the document's metadata tags and the metadata taxonomy. The taxonomy is used to create a tree structure and the metadata tags are added to this as leaf nodes. Each tag corresponds to one row in the central grid. The iconic document overview is a miniature image of the document laid out horizontally along the top of the visualisation. Each paragraph in this overview lines up with one of the columns in the grid below. The central grid shows which metadata tags have been applied to which paragraphs in the document; each column represents a paragraph and each row represents a tag. Hence, to show that a tag has been applied to a particular paragraph the cell where the appropriate column and row meet is shaded. The colour of the cell is determined by the value of the applicability attribute given to the tag when it was applied to this paragraph; the higher the applicability the darker the colour (a legend for this scale is always visible in the top left hand corner). The metadata tags and taxonomy, iconic document overview and grid, are thus brought together to produce a visual overview of the document.

The user can query the visualisation with their mouse through a combination of dynamic querying and brushing [3]. When the mouse is placed over a cell in the grid, all the tags in the column are highlighted; hence the paragraph's content is described. By looking along the row of an interesting tag, the user can see where and to what extent that topic comes up in the document. So by moving the mouse in the central grid the user can explore what different parts of the document are about, and where in the document particular topics are covered.

The user can also query the document by clicking on the any cell in the grid. Clicking on the grid will produce a query that is sent to the servlet formulated with the metadata and paragraph corresponding to the selected cell. The servlet will produce an HTML version of the document in which every paragraph tagged with the selected metadata is highlighted using a bold font. The browser will display paragraph selected; thereby offering details on demand [3].

3 Evaluation

An experiment was undertaken to see whether Gridvis offered a performance advantage for goal-directed search tasks.

3.1 Method

The documents selected were a set of five reports written by an analytical research team within Sainsbury's. These are substantial papers, which may inform business choices on specific supply chain issues for the foreseeable future. Five of these documents were used in this experiment; these were tagged with metadata, and paragraph numbers.

Three information needs for each document were elicited from the original audience of the documents. One to two people were interviewed for each document (corporate restructuring meant this was the best that could be done), and were asked to say what they had been looking for, or would look for now, when they read the document. This process was undertaken to ensure the information needs used would be as representative as possible of what Gridvis would be used for in a real work situation.

The relevance of paragraphs in the document to each information need was established using three judges and an adjudicator. In each case one of the judges was the documents author, and the other two had been subjects in the experiment described below. They were asked to read through the document carefully to locate every relevant paragraph. They were asked to classify the relevance of each paragraph as one of 'contains key information', 'contains relevant information', and 'contains related information'. An adjudication session, including three of the judges and the researcher as an adjudicator, was used to decide the relevance of paragraphs where the original judgements showed large disagreements. Where the differences were small, a mean score was calculated. This process resulted in a single set of relevance judgments with which the retrieval performance could be measured.

Twelve participants were found who were employees of Sainsbury's and had not read the documents, yet claimed to be familiar with the concepts in them. Each had the visualisation explained to them and were asked to

familiarise themselves with it by attempting to satisfy three information needs as quickly as possible with the first document. The subjects were then asked to repeat this task with four other documents. Two of the documents were presented with Gridvis; the other two were displayed alone in a browser window. When using Gridvis, the subjects were asked to only make minimal use of the document window's scroll bar and thus rely on Gridvis for navigation and search.

The detail of users' interactions were preserved through a continuous screen recording of the entire experimental session. Every document query action (i.e. when the participants clicked on a cell in the grid) was automatically recorded in a text file and the relevance of paragraphs found were noted by the subjects. They were asked to record the paragraph numbers of germane paragraphs and to classify the paragraph's relevance. A paragraph could be given three levels of relevance, as described above.

3.2 Results and discussion

A quantitative analysis of anything but large performance differences is not a realistic part of academic research in a corporate environment. For the effect size observed (d=0.55, a medium effect size) we would require a sample size of 126 to perform a sufficiently powerful two sample t-test [4]. Nevertheless in the hope of large performance differences, data from this experiment was used to calculate average search times, as well as scores for precision (the proportion of answers that were correct) and recall (the proportion of all the possible correct answers obtained). The medium effect size and large variance observed precluded their being any significant difference between performance with and without Gridvis using only 12 subjects.

A qualitative analysis of the data was therefore undertaken to determine why Gridvis did not produce a large performance improvement and what improvements might be possible. This analysis started by looking at what strategies the subjects might be using to make queries with Gridvis.

The results of an earlier task analysis of Gridvis use were used to focus the analysis of strategy usage. The task analysis divided document query production into two components. Firstly, the user has to find a relevant metadata tag, then they must to select from amongst the paragraphs relevant to this tag. The task analysis also suggested a number of strategies that might be used to make these selections.

This initial set of strategies was used to classify each tag and paragraph selection action of the 12 subjects. The classification was made on the basis of inspecting video and text log files of the session. During analysis of the first few subjects the classification selection was refined until they could account for 95% of tag selections and 88% of paragraph selections. The strategies used in this classification will only be listed briefly, with some being explained as they are raised later discussion. The tag selection strategies were: keyword mapping, synonym keyword mapping, related concept keyword mapping, and narrative keyword mapping. The paragraph selection strategies were: 'left-to-right', 'applicability' only, 'co-occurrence' only and finally 'applicability and co-occurrence'.

The frequency of usage of different tag selection strategies was examined first. Subjects were found to use a range of strategies, with no one strategy being unduly neglected. The only problem observed here was a tendency for subjects to overlook some relevant tags as evidenced by the of use a mean of 3 out of the mean 6 tags relevant to each information need.

However, behaviour seemed more problematic in the use of paragraph selection strategies to pick paragraphs related to a tag of interest. Here the users were found to seldom use strategies that capitalise on information given by the visualisation. Only a third of tags were queried using the 'applicability only' strategy. This strategy involves inspecting paragraphs in an order based on applicability information (i.e. cell shading colour), which indicates how relevant a tag is to a paragraph. Additionally, users seemed to make little use of the 'co-occurrence only' strategy. Co-occurrence is essentially information about what other tags have been applied to a paragraph of interest. This information is important since it can serve to more fully describe a paragraph's content and disambiguate metadata tags. The analysis revealed that co-occurrence seemed to be used to select paragraphs only for a fifth of the tags inspected. Hence, in the case of paragraph selection, the information offered by the visualisation was little used.

This apparent under-utilisation of the information provided by Gridvis suggests performance improvements would be realised if the information were used. The validity of this suggestion depends on the assumption that making full use of information about applicability and co-occurrence will lead to better performance. An initial test of this assumption was undertaken by exhaustively applying the tag and selection strategies listed above.

Recall (proportion of total relevant paragraphs located) was increased to threefold of that achieved by subjects in across conditions by exhaustive application of tag selection strategies. It was also found that the resulting

decrease in precision (the proportion of located paragraphs that were relevant) could be offset using the paragraph selection strategies. These strategies were used to rank the set of paragraphs related to the tags identified with the tag selection strategies. The top-ranked half of this set of paragraphs had a recall and precision about 60% higher then the bottom-ranked half. Hence, using the tag and paragraph selection strategies can be used to increase precision and recall.

If making more use of the information given by the visualisation would have improved performance, why did users not do so? Lack of familiarly and expertise might be one reason. A more important problem is that the cost of using this information might be too high. This seems particularly true for co-occurrence information. To use this, the user must read the metadata tags applied to each paragraph. The tags for each paragraph must then be held in memory while their relative relevance is assessed. Moreover, the application design makes using co-occurrence information particularly difficult and time consuming in some circumstances. For example, if there is more then one interesting tag, the user might want to find paragraphs relevant to them all. This requires the user to scan along the row of each interesting tag to find coloured cells. Upon arriving at each of these, the user must then scan vertically to the check if the cells relating to the other tags of interest are coloured in.

New functionality is being added to Gridvis, to address these problems. The co-occurrence strategy can be partially automated if the user selects tags of interest and Gridvis indicates the number of tags that co-occur on each paragraph. This will also allow users to find paragraphs tagged with all the metadata of interest.

Another possible factor inhibiting the use of applicability and co-occurrence information is Gridvis' voracious appetite for screen real estate. With longer documents the users have to scroll horizontally to see the whole grid. This makes it impossible to see every paragraph relevant to a tag and therefore confounds attempts to select the most appropriate paragraph. This problem is being addressed by a new implementation that dynamically redraws the grid at varying levels of detail to only show in full the parts of the document relevant to the currently selected tag(s).

3.3 The problems of empirical evaluation in a corporate environment

Robust quantitative empirical evaluation using controlled experimentation often depends, as it has with Gridvis, on having a reasonably large sample size. This presents an insurmountable problem to the academic researcher working in a corporate environment.

Gridvis was created for use in a corporate environment. Its design has therefore evolved around the corporate material (documents, metadata etc) it uses. These materials, however, mean little to the uninitiated. Also, for the evaluation to meaningfully relate to use in the work situation, representative information needs were elicited from Sainsbury's staff. Only people with the appropriate background knowledge will fully understand these information needs.

Therefore for Gridvis to be evaluated, subjects must be found who understand the documents, metadata taxonomy, metadata and information needs used. The implication is that participants must be Sainsbury's employees from the parts of the business close to the documents used. This means that obtaining more then the twelve willing subjects found for this experiment would be extremely challenging. Obtaining the 126 required for a sufficiently powerful statistical test is virtually impossible. Another problem is that one can only ask these subjects to devote a small portion of their time to an evaluation; an hour at most. This did not seem to be enough time for users to become truly confident with the application before doing the experiment. Consequently, this kind of empirical evaluation can only test the performance of novice users.

An additional limitation of this approach is that the already prohibitive size and cost of experiments dramatically increases with the number of questions examined. It is therefore unsuitable for answering the many detailed question needed to inform application design and improvement.

An alternative approach, perhaps more suited to evaluation in a corporate setting is performance modelling. We are currently using quantitative optimisation models of information seeking developed at Xerox PARC [5] to predict the best performance achievable with the interface given a set of well-defined strategies. This 'best performance' can then be compared to a control condition i.e. performance for the same task with paper documents where people are already expert performers. Such comparisons needs much smaller sample sizes of around 25 subjects [4]. Another advantage is that many different task parameters (e.g. document length, query difficulty) and design possibilities (e.g. improving the speed of certain operations) can be examined more cheaply and more powerfully.

4 Conclusion

This paper has described Gridvis, a tool for rapid goal-

directed retrieval of information within a document. It uses visualisation to enable humans to make direct use of metadata that describes content at a paragraph level. The visualisation is an overview of document content which allows the production of simple dynamic queries.

Empirical evaluation was used to examine whether Gridvis increases the efficiency of information location. Subjects were asked to satisfy a set of information needs using online documents with, and without, the aid of Gridvis. Quantitative analysis did not reveal a significant numerical improvement in performance when using Gridvis. However, qualitative analysis suggested that subjects were not making full use of the metadata. This result helped to identify several problems with the Gridvis interface, which are currently being tackled through a set of improvements.

We have encountered important constraints on what methodology is used for reliable evaluation of visualization performance in a corporate context. The research must be sensitive to limits in the amount of business resources it can expect. It will rarely be tractable to obtain the large sample sizes of staff often needed for controlled experiments. For this reason it is suggested that a model-based approach provides an evaluation methodology better suited to such industry based academic work.

5 Acknowledgments

This research was sponsored by J Sainsbury plc and undertaken within the Postgraduate Training Partnership established between Sira Ltd and University College London. Postgraduate training partnerships are a joint initiative of the DTI and the EPSRC.

6 References

[1] A. Adler, A. Gujar, B. L. Harrison, K. O'Hara and A. Sellen, A diary study of work-related reading: design implications for digital reading devices; Conference proceedings on Human factors in computing systems , 1998, Pages 241 - 248

[2] V. R, Benjamins, D. Fensel and S. Decker, $(KA)^2$: building ontologies for the internet: a mid-term report, International journal of Human-Computer Studies (1999) 51, 687-712

[3] S. K.Card, J. D. Mackinlay, & B. Shneiderman (Eds) (1999) Information visualisation: Using vision to think, Morgan Kaufmann, p 223

[4] D. C. Howell, Statistical Methods for Psychology (4^{th} edition) Duxbuy 1997

[5] P. Pirolli and S. K. Card (1999). Information Foraging. Psychological Review 106(4): 643-675.

[6] M. Uschold (1996) Building Ontologies: Towards a Unified Methodology, http://www.aiai.ed.ac.uk/~entprise/enterprise/ontology.html#papers - abridged version appears in Proceedings of the Workshop on Ontological Engineering in conjunction with ECAI 96

[7] M.Uschold, M.King, S.Moralee and Y.Zorgios (1998) *The Enterprise Ontology* The Knowledge Engineering Review , Vol. 13, Special Issue on Putting Ontologies to Use (eds. M. Uschold and A. Tate).

[8] N. A. Van House, User Needs Assessment and Evaluation for the UC Berkeley Electronic Environmental Library Project, Digital Libraries '95: The 2^{nd} Conf. on the Theory and Practice of Dig. Lib., June '95, Austin, US