MS Read: User Modeling in the Web Environment

Nataša Milić-Frayling
Microsoft Research Ltd.
St. George House, 1 Guildhall Street
Cambridge CB2 3NH
Tel.: +44 (0)1223 744 744
natasamf@microsoft.com

Ralph Sommerer
Microsoft Research Ltd.
St. George House, 1 Guildhall Street
Cambridge CB2 3NH
Tel.: +44 (0)1223 744 744
som@microsoft.com

ABSTRACT

MS Read is a prototype application implemented as an extension of the Web Browser that creates an evolving model of the user's topic of interest. It uses that model to analyze documents that are accessed while searching and browsing the Web. In the presented version of MS Read the model is used to highlight topic related terminology in the documents.

MS Read model of the user need is created by applying natural language processing to search queries captured within the Browser and to topic descriptions explicitly provided by the user while browsing and reading documents. It is semantically enhanced using linguistic and custom knowledge resources.

1. INTRODUCTION

In the WWW environment, the *separation of information providing services*, such as search, *from document delivery* is evident. The most crucial consequence of the separation between document delivery and search is the loss of (search) context that could be used to analyze delivered documents. For that reason many beneficial features such as *query specific* analyses and presentation of documents are rare or absent from the Web applications and services.

Our analysis of the Web Search scenarios points to the Web Browser as an essential facilitator of the communication between the Web Search Service, the user, and the information hosting Web servers. Thus we implemented MS Read, an enhancement of the Browser that accomplishes two main objectives: (1) captures the context of search, independent from any particular search engine, and applies it to documents accessed via the Browser while searching or browsing, and (2) enables personalization of the document analysis through the enhancement of the context model with the user and task specific knowledge resources.

2. DESCRIPTION

MS Read's core functionality comprises:

- Module for analyzing document format (e.g., HTML parser)
- Module for the linguistic analysis of text (e.g., linguistic processing of the user's topic descriptions)
- Linguistic and knowledge resources for enhancing the context.

As the user communicates a request for information via the Browser to a Web information providing service, MS Read analyzes the HTML and the events associated with the Web page to identify the user's input and decide whether it is search related. This user input is passed onto the linguistic module based on MS NLPWin software [1] that analyzes the text and creates an elaborate linguistic representation of the user's request: identifies concepts (phrases and single word terms without pre-modifiers) and keywords of various linguistic characteristics (head nouns, verbs, etc.). More details are presented in [2].

Furthermore, MS Read GUI allows the user to search through the viewed documents by providing additional topic description to MS Read if needed. Finally, it uses available language and knowledge resources to further enhance the context model.

In order to illustrate the potential of this approach, we implemented a context sensitive highlighting of documents. Benefiting from the extensive context modeling on the topic side, MS Read performs effective highlighting by applying pattern matching of the context terminology to the viewed documents.

We should point out that MS Read is independent from any particular search engine. It provides assistance in reviewing documents both while searching and browsing. Finally, it is extendible: it can incorporate as much information about the task and the user as the client's computing power allows.

3. ACKNOWLEDGMENTS

Many thanks to Robert Tucker who worked on the prototype and Libby Hanna who conducted a user study of an early prototype.

4. REFERENCES

[1] Jensen, K., Heidorn, G.E., Richardson, S.D., eds. *Natural Language Processing: The PLNLP Approach*. Kluwer Academic Publishers, 1993.

[2] Milic-Frayling, N., Sommerer, R., *MS Read: Context Sensitive Document Analysis in the WWW Environment*, Microsoft Technical Report: MSR-TR-2001-63, June 2001, http://research.microsoft.com/scripts/pubs/trpub.asp.

Copyright is held by the author/owner.
SIGIR'01, September 9-12, 2001, New Orleans, Louisiana.
ACM 1-58113-331-6/01/0009.