A PERSONALIZED RECOMMENDER SYSTEM FOR WRITING IN THE INTERNET AGE

M. C. Puerta Melguizo[1], O. Muñoz Ramos[1], T. Bogers[2], L. Boves[1] and A. van den Bosch[2]

[1]*Department of Language and Speech, Radboud University, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands*

[2]*ILK/Language and Information Science, Tilburg University, P.O. Box 90153 NL 5000, LE Tilburg, The Netherlands*

Keywords: Proactive recommender systems, writing stages, information seeking, long-term memory.

Abstract: With the advent of Internet, writing and finding information to plan and structure the text have become increasingly intertwined. We think that it is necessary to develop systems able to support the task of finding relevant information, without interfering with the writing process. The Proactive Recommender System À Propos is being developed in order to support writers in finding relevant information during writing. We present our research findings and raise the question whether the tendency to interleave (re)search and writing implies a need for developing more comprehensive models of the cognitive processes involved in writing scientific and policy papers.

1 INTRODUCTION

Writing professional documents (e.g. scientific papers, user manuals, etc) is complex. The most widely influential model of writing is the one proposed by Hayes and Flower (1980). Although in this model the processes considered are the ones involved in writing with pen and paper, virtually all software systems that have been design seem to build on the concepts developed in this model. We will start by introducing the model of Hayes and Flower and how the use of computers and the internet has changed the way we write. We will finish by presenting research we are performing in order to develop a Proactive Recommender System: À Propos. Our research is based on the conviction that in order to design better tools for writing it is important to understand the cognitive processes involved in writing and searching information.

2 THE COGNITIVE PROCESSES OF WRITING

According to Hayes and Flower (1980), writing happens in three stages: Planning, Translating, and Reviewing. During *Planning* ideas are generated and arranged into a coherent structure. Planning involves retrieving domain knowledge from the writer's

Long-Term memory (LTM). During *translating* writer's plans are transformed into sentences. In the *Reviewing* stage the writer evaluates the relation between the text written so far and the linguistic, semantic and pragmatic aspects that would best serve the writing goal. Reviewing involves reading and editing errors or weaknesses in the text. The *task environment* includes everything outside the writers' mind that can influence the writing task including the text produced so far and the so called rhetorical problem (the writing assignment, the specification of topic and the audience). In the *writer's LTM* are stored the writer's knowledge about the topic, the knowledge of sources based on literature search, the writing plans and the knowledge about the audience who will read the work.

Hayes (1996) extended the model and emphasized the role of working memory, as well as socio-cultural and motivational aspects in writing. Furthermore, the task environment is divided into social and physical contexts. According to Hayes the *social environment* needs to be considered because writing is a social activity, and consequently, the way a text is written is affected by several cultural conventions and the audience it is meant for. In *the physical environment*, the *composing medium* or tool used to write has been added to the text produced so far. Actually, variations in the medium seem to lead to differences in the way people carry out the writing task. For example, Haas (1996) found that writers

tend to plan more and review at a more general level when they write on paper than when using a word processor. These results suggest that the introduction of computer tools seems to force users to change the processes they use. However, still a lot of research needs to be done in order to explore how the composing medium affects the writing process.

3 WRITING IN THE INTERNET AGE

Current models of writing assume that knowledge about the topic of the text is mainly stored in the writer's neural LTM. The reality of writing professional texts however shows that writers almost invariably need to look for additional external information while writing. And with the advent of the Internet more frequently than ever, writing is now interleaved with searching for information. Yet, seeking for information is difficult and time consuming. Keyword-based search is still inefficient and relevant information may be missed. Also considerable time is spent interacting with low-precision search engines. Consequently, the time in which the author is away from creating the document can have a negative impact on the total time spent and on the quality of the text. Furthermore, we question whether continuously switching between writing and searching is efficient, and whether it tends to result in the best possible quality of the texts. Finally, we think it is necessary to design tools that support writing and help users to retrieve relevant information.

4 À PROPOS

A Proactive Recommendation System (PRS) relieve authors from explicit search and switching between applications by means of searching information accurately and recommending this information in a proactive manner. For example, Watson (Budzik and Hammond, 1999) performs automatic Web searches based on text being written or read. A problem with current PRSs is that they are developed as search tools and do not take into account the specific characteristics of the writing task.

Our goal is to develop a PRS for writers in a professional environment: À Propos. The architecture is based on a client-server architecture. The client runs on the user's computer and monitors user's activity constantly. À Propos proactively submits queries based on the user and group profiles in combination with what the user is currently typing or reading. The server consults the relevant information sources, and returns the search results to the client. A more detailed description of the system's architecture can be found in (Puerta Melguizo et al., 2007a) where the role of the different components of the system such as observers, filters and gatekeepers is explained. In the User Interface the results of the search are presented in a semi-transparent window located in the bottom right of the screen (see Figure 1). The window contains URLs related to what the user is typing. As the user moves the cursor over the references, the URLs become fully visible and active. On clicking the required URL, the user accesses the corresponding paper from the digital library. The information in the window changes depending upon the text that is being input and new queries that are created. To develop À Propos two main issues are being researched. First, in order to present highly relevant information, appropriate filtering techniques need to be developed. Second, procedures to identify the different writing stages and related information needs must be created in order to design an appropriate user's interface. The researches performed for both issues are discussed below.
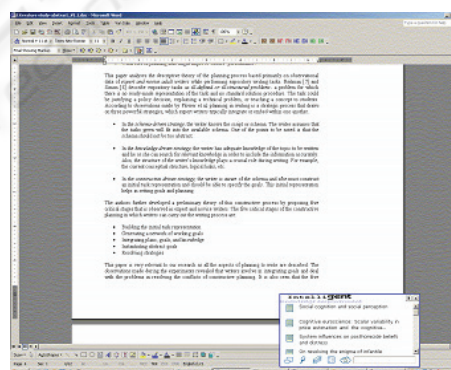

Figure 1: The user's interface.

4.1 Selecting and Presenting Relevant Information

The acceptance of any PRS hinges on the relevance and accuracy of the suggested information. Quality recommendations should be both on topic and personalized. To increase the topicality of the recommendations one can use detailed personalized taxonomies integrated in an easily expandable, yet robust IR model to retrieve the initial list of documents. We are investigating personalization on user and group level.

4.1.1 User Personalization

When personalizing results, we consider the user's interests and expertise. From these data we build a profile of terms important for the user which is used to re-rank the initial recommendations and suggestions with more matching profile terms get promoted to the top of the list. Three different sources of information are considered for inclusion in the user profile: past selections from the list of recommendations, user's past documents, and the PRS also allows users to enter informational queries manually.

4.1.2 Group Personalization

À Propos aims to perform group personalization by identifying the expertise in different topics of the members of a group. The user's own documents and profile are seeing as an expertise fingerprint of that user. We can then use taxonomies (e.g. the ACM hierarchy) to represent the hierarchy of topics for which we want to quantify a group members' expertise. By collecting an adequate number of documents for each topic we can extract the representative terms and construct topic fingerprints. The next step is to match these topic fingerprints with the user's expertise profile by calculating the term overlap. This way we can calculate the expertise of each group member on the different topic areas and also find out which group members are experts in the topic of the user's active document. Knowledge of the distribution of expertise over the group can then be used for personalization. For instance, the recommendation of a document by an expert on the topic should be considered as more reliable and have a significant influence on the final re-ranking (Bogers and Van den Bosch, 2006). Group personalization could also be used to recommend documents that were not even in the initial recommendation list. Expertise fingerprints can also be compared to each other and used to suggest related topics to the user to provide for a more serendipitous experience. Our experience suggests that serendipity is especially important in the earliest phases of planning.

4.2 The Problem of Interrupting the Stages of Writing

One problem with presenting proactive information is that it can interrupt the ongoing writing task. The interruption can also be more disturbing and distracting in specific stages of the writing process.

Consequently, the effects of interruptions during different writing stages need to be considered. Deshpande et al., (2006) found that writers need to look for extra information especially during planning and reviewing. Consequently, we decided to study the effects of presenting proactive information during these stages (Puerta Melguizo et al., 2007b)

4.2.1 Presenting Proactive Information during Planning Tasks

To simulate the stage of Planning, participants were told that to write essays, they had to start by writing an outline of the major points and order in which they would be introduced in the essay. The writing outline was the planning task. Participants wrote the planning outlines: 1) without PRS and no option of looking for extra information, 2) without PRS and the option of getting information by actively searching information in the Web, 3) with presentation of proactive relevant information by our PRS, and 4) with presentation of non-relevant information by our PRS

The PRS did not seriously impair time performance. Furthermore, when relevant information was presented proactively, the quality of the writing plan was significantly better and participants introduced more information than in the other conditions. The results of this experiment also show that active search initiated by the user resulted in a lower quality of the information found and a worse written text.

4.2.2 Presenting Proactive Information during Reviewing Tasks

Participants performed two editing tasks: spelling corrections and filling in factual information in the text. Participant performed both reviewing tasks under three conditions: 1) without PRS and the option of getting information by actively searching information in the Web, 2) with presentation of proactive relevant information by our PRS 3) with presentation of non-relevant information by our PRS.

Again, the presentation of proactive information did not impair time performance. Furthermore, the time spent in looking for new relevant information was shorter when the PRS presented relevant information than when participants searched for the information actively. The information seeking time was even longer when non-relevant information was presented proactively. In this case, after assessing that the information by the PRS could not help in

completing the editing task, participants started an active search. This result emphasizes the importance of developing appropriate search profiles and filters as described above. Finally, the quality of the editing tasks was also significantly better when proactive relevant information was presented showing once more, that active search initiated by a user is less effective.

4.3 An External Long-Term Memory

Virtually all writing research has been conducted in settings in which the LTM from which participants could 'get information' was limited to their own brain. However, the advent of Internet is already affecting the way people consider and use LTM and now is becoming more important to know how to find information than to memorize information in the first place. However, accessing information in the Internet is not without problems. Knowing less, while searching more makes more difficult to assess the importance of found information and to integrate it in a coherent framework. A PRS could be able to support the decisions about the relevance of the results returned from a query and be used as an addition to the writer's neural LTM. Furthermore, we think it is neccesary to develop a new model of cognitive writing processes in which the external LTM that the WWW and other databases conforms, needs to be included as an important part of the physical environment.

5 CONCLUSIONS

In this paper we presented the PRS À propos. This system is in development and aims at supporting writers in the difficult task of finding appropriate relevant information during writing.

First, we present the efforts we are performing in order to develop adequate group and personal profiles that make sure the information presented by the system is relevant to the writer and to the specific piece of text is being written. We also describe the studies we performed in order to explore the effects of presenting proactive relevant information when writers are planning and reviewing text. From our experiments we could conclude that the user's interface of the PRS does not negatively interrupt the task of writing. And even more important, when relevant information is presented, the quality of the writing text significantly improves in comparison with the situations in which the user actively seeks for

information. Furthermore, the results of our experiments with proactive presentation of information suggest that professionals are willing to accept unsolicited pop-up windows and similar interrupts if the information that they are alerted to by those interrupts is relevant for the completion of their (writing) task.

REFERENCES

Bogers, T. and Van den Bosch, A., 2006. Authoritative Re-ranking of Search Results. In *Proceedings of the 28th European Conference on Information Retrieval (ECIR 2006), vol. 3936 of Lecture Notes on Computer Science*, pp. 519-522. Springer Verlag, April 2006.

Budzik, J. and Hammond, K., 1999. Watson: Anticipating and Contextualizing Information Needs. In *Proceedings of the 62nd Annual Meeting American. Society for Information Science,* 727-740.

Deshpande, A., Boves, L. and Puerta Melguizo, M.C., 2006. À propos: Pro-active personalization for professional document writing. In *SigWriting, 10th International Conference of the EARLI Special Interest Group on writing.* Antwerp, Belgium.

Haas, C., 1996. *Writing Technology Studies on the Materiality of literacy*. Lawrence Erlbaum Associates, Hillsdale New Jersey.

Hayes, J.R., 1996. A new framework for understanding cognition and affect in writing. In C.M. Levy and S.E. Ransdell (Eds.). *The science of writing: Theories, methods, individual differences, and applications* (pp. 76-97). Lawrence Erlbaum Associates, Hillsdale New Jersey.

Hayes, J.R. and Flower, L.S., 1980. Identifying the organization of writing processes. In L.W. Gregg and E.R. Steinberg (Eds.). *Cognitive processes in writing* (p. 3-30). Lawrence Erlbaum Associates, Hillsdale New Jersey.

Puerta Melguizo, M. C., Bogers, T., Deshpande, A., Boves, L., and Van den Bosch, A., 2007a. What a proactive recommendation system needs - relevance, non-intrusiveness, and a new long-term memory. In *Proceedings of the ICEIS: 9th international conference on enterprise information systems*, 86-98.

Puerta Melguizo, M.C., Boves, L, Deshpande, A., and Muñoz Ramos, O., 2007b. A Proactive Recommendation System for Writing: Helping without Disrupting. In W-P. Brinkman, D-H. Ham and W. Wong (Eds.). *ECCE 2007: European Conference on Cognitive Ergonomics*, 89-95.