Personalized Environment for Skimming Documents

Tessai Hayama

School of Knowledge Science,
Japan Advanced Institute of Science and Technology
March 2003

Keywords: knowledge sharing, text summarization, collaborate model, sentences extraction algorithm, information visualization

The purpose of this paper is developing environment for skimming documents. Reading document is important for our creative activities (for example, study activity, research activity and etc). Recently, there is much number of on-line documents in our surroundings, and the readability of document from computer screen is an issue that has increased in importance with the increasing frequency and range of computer screen use. We, however, prefer on paper to on screen in reading documents. For this reason, it is poor at skimming documents, which is our reading style, from computer screen. The methods, to support speed and comprehension in reading by the advantage of a computer, are interfaces for reading, text summarization and etc. Interfaces using visualized effect help user's reading documents (for example, liner effect, fisheye effect, overview+detail effect and etc). Automatic text summarization is useful to understanding contents of documents immediately, it, however, is imperfect. In this paper, we develop personalized environment for skimming documents, which include two approaches: 1) skimming support system that has interface in consideration of the fault of reading documents from computer screen and advantage of computer, 2) recommend system which recommend personalized summaries. 'Skimming' is defined as proceeding at a rate three to four times faster than normal reading "in order to grasp a general sense of the content or to retain only the main points", and target documents of this paper are Japanese theses.

The developed environment for skimming documents is server/client system, client

system is skimming support system and server system is recommending summary system. It is effective for skimming documents to support the important position and the present reading position in memory. The interface of the skimming support system, thereby, is using effect combined fisheye and overview-detail. The focus points of fisheye effect are sentences extracted by using the original sentences extraction algorithm. This algorithm is following; calculate word weight by distribute value between section (or subsection) and segments, calculate sentence weight by add word weight in the sentence, extract high score sentences in 20%. The overview interface of this system is automatic generation of tables of contents. The purpose of overview interface, generally, is to support for users to memory reading present position. Users prefer paging to scrolling in reading document from screen, because the whole long document cannot be displayed for physical space of screen. The detail shows a segment unit of long documents in each page. Our using segmentation method is the method based on the TextTiling algorithm. Another developed system is personalized recommend system of summaries by using user's profile, which create different sentences selected by user in the skimming support system, because each users has different reading points. After using, the skimming support system sends and saves user's log, number of selected sentences, to sever. The recommend system calculates similar of theses and sentence numbers from other user's log, and recommend summaries created by most similar user. The similarity is calculated by using Jaccard coefficient.

In evaluation experiences it evaluated three points, which are the sentences extraction method, support to skim documents from computer screen, and recommended summaries. In the evaluation experience of the sentences extraction method, this method was more effective value than the other method and F-measure of the average value is 0.48. In the evaluation experience of support to skim documents from computer screen, past record is that the effective reading rate of reading from paper is 41% higher than that from computer screen, skimming documents by using this system from computer screen, however, are the same effective as that from paper. In the experience of recommended summaries, correlation of summaries between the users chose and the system chose was discovered. It, also, discovered correlation of user's log of the skimming support system and user understanding.