Abstract

The overall purpose of the present thesis is to investigate, if automatic assigned indexing methods can improve professional users' access to work-based documents in the domain of e-government. The problem is investigated by means of a case study in the Danish tax authorities SKAT. An experimental comparative test was designed on the basis of a preceding domain study, clarifying the seeking behaviour in e-government.

The introduction of e-government has arisen from a desire for effectiveness, efficiency and greater transparency in public administration. Today public-sector employees commonly carry out manual indexing of government documents. With the thesis we want to investigate if automatic indexing can replace, and perhaps even improve, the current manual procedures to be able to support efficiency and effectiveness.

An employee perspective guides the thesis. That involves a user group with great knowledge of the topic they are working with. In contrast to citizens and other e-government stakeholders, not much is known about the seeking behaviour of employees in the domain. In addition the introduction of e-government is expected to change employees' work tasks, and with that their information needs. That calls for an investigation of the present information seeking behaviour of e-government employees. In the thesis this is done by means of a domain study. The study is based on a questionnaire distributed to employees in SKAT and subsequent focus group interviews. The domain study shows that the employees use a number of primarily online information sources to solve their work tasks. The sources are used frequently. The employees primarily have verificative and conscious topical information needs. Besides that they are experienced information searchers requesting more extensive metadata in the system forming the basis of the search test: their intranet.

The knowledge gained from the domain study was incorporated into the search test design. The test was an experimental test comparing automatic extracted indexing (free text indexing) and automatic assigned indexing (categorization). In the assigned indexing a domain specific taxonomy formed the basis of the categories. The test system was a prototype of a future version of SKATs intranet. 32 test persons carried out searches with the two indexing types in two separate systems in experimental sense. 3 simulated search tasks and 1 genuine search task guided the searches. The the

simulated search tasks were designed in accordance with the findings from the domain study regarding the information needs of the employees. The test showed that the two automatic types of indexing are useful to the employees in their own way. At a general level extracted indexing had the best performance measured in terms of the average number of terms and concepts in queries, in terms of the number of sessions with reformulations, and in terms of the number of reformulations in sessions. This showed that the system with categorization demanded more from the test persons in comparison to the free text indexing.

It turned out that the test persons had difficulties using the categorization in some respects. Thus it was not relevant to them, if they retrieved a highly relevant document with a high rank order before using the categorization. They did not find it relevant either, if they retrieved very few results by the initial search. In those cases it was easier for them to manually go through the results. In contrast the categorization was helpful in identifying new facets of a search task and in suggesting new search terms in reformulations. For future e-government indexing guidelines this resulted in the recommendation that both assigned and extracted indexing should be represented as search facilitators, as they support their own aspects of the information needs arising for employees in e-government.

The thesis contributes by providing new insights into the information seeking behavior of employees in e-government and the way in which this behavior can be supported by automatic indexing.