Bar-Ilan University (RA)
Faculty of Humanities
Library of Information science

אוניברסיטת בר-אילן (ע"ר)
הפקולטה למדעי הרוח
הספריה ללימודי מידע

The experience of automatic classification of personal information a simulation research / Edva Lotan

Abstract

This research explores the field of Personal Information Management, which focuses on the challenges facing computer users when dealing with the classification of their personal information and retrieving it for various purposes. These challenges become more and more complex with the increase in the amounts of personal information, requiring users to invest time and make a cognitive effort in order to organize their information so that it will serve them best.

Classifying the information in the computer can be done in different ways. A common method is placing files in a hierarchal folder structure. Classifying the information, including maintenance of the folders, requires a cognitive effort and when done frequently or in large quantities, becomes a tiring and expensive endeavor.

Among other solutions used to improve the organization of personal information, there's an examination of the option for an automation solution: automated software programs which operate independently without the user's involvement, that are intended to assist him in his work. These software programs are called "agents" and they are a subject of research in the field of artificial intelligence.

An agent is a system that operates intelligently according to changing goals and circumstances, learns from experience and makes suitable decisions under the limits of its sensory abilities and computational power. In light of the advancement in technology, it is not unlikely that in

_____

Library of Information Science   הספריה ללימודי מידע
Bar-Ilan University, Ramat-Gan, Israel   52900 אוניברסיטת בר-אילן, רמת גן, ישראל
/library www.is.biu.ac.ilWebsite: Email: Ruthi.Tshop@mail.biu.ac.il   Tel. 972-3-5318163 .טל

Bar-Ilan University (RA)
Faculty of Humanities
Library of Information science

אוניברסיטת בר-אילן (ע"ר)
הפקולטה למדעי הרוח
הספריה ללימודי מידע

the future agents could classify documents at an accuracy level resembling that of a human assistant. Yet the technology is not the only relevant aspect: another consideration must be the way users react to the utilization of such software in regards to their user experience, especially their feelings of trust, control, security and comfort.

This research aims to answer two main questions: first, what is the user experience when using an automated solution of personal information classification? And secondly, what are the attitudes towards the idea of using a software program for automatic classification of personal information? To answer these questions, an experiment was conducted, which included a human simulation of automatic classification of personal information, followed by interview that was analyzed qualitatively. Also, a questionnaire regarding the use of such a software was distributed..

The simulation experiment included 20 test subjects who manage personal information in files. During the experiment, the researcher learned the classification methods of each test subject. Then, the test subjects gathered files they created, located or received in the duration of a week and the researcher classified them according to the rules described to her by the test subject. Finally, a semi structured interview was conducted, during which the test subjects described the quality of the classification performed by the researcher, commented on the user experience of the simulation experiment and shared their position regarding the use of an automatic classification software. The questionnaire was distributed among 219 participants from the same sampling. It included 9 statements, each expressing an opinion regarding personal information management in general and specifically regarding automatic classification.

_____

Library of Information Science
Bar-Ilan University, Ramat-Gan, Israel
/library www.is.biu.ac.ilWebsite: Email: Ruthi.Tshop@mail.biu.ac.il   Tel. 972-3-5318163

הספריה ללימודי מידע
אוניברסיטת בר-אילן, רמת גן, ישראל 52900
טל. 972-3-5318163

Bar-Ilan University (RA)
Faculty of Humanities
Library of Information science

אוניברסיטת בר-אילן (ע"ר)
הפקולטה למדעי הרוח
הספריה ללימודי מידע

Relevant categories regarding user experience over the course of the simulation experiment were identified during the analysis of the interviews, and were grouped into parent categories according to subject and context (for example, aspects of control, trust, comfort, etc.) and whether the statements were of a positive or a negative nature. The quality of files classification by the researcher was measured according to the score given by each test subject and also according to whether the test subject was aware of mistakes made during the classification. A similar analysis was performed to determine the participants' position towards using an fictional automated classification software. Also, the findings of the questionnaire were analyzed using descriptive statistics tools.

Overall, the test subjects' user experience was good. The average of positive statements per test subject was 67%, with a standard deviation of 30%. According to the researcher's impression, 35% of the test subjects had a good experience and 35% had a pretty good experience during the simulation experiment. For the majority of test subjects, 16 out of 20, the experience was mixed and they expressed both positive and negative feelings.

Prominent categories of a positive nature were: a feeling of trust toward the classification (12 test subjects); feeling the classification preserved their work habits (10 test subjects); a feeling of control (9 test subjects); a feeling of value (8 test subjects). Prominent categories of a negative nature were: discomfort regarding participation in the experiment (7 test subjects); a feeling of invasion of personal space (5 test subjects); a lack of value (5 test subjects). No unequivocal connection was found between the findings of user experience and the quality of file classification.

_____

Library of Information Science   הספריה ללימודי מידע
Bar-Ilan University, Ramat-Gan, Israel   52900 אוניברסיטת בר-אילן, רמת גן, ישראל
/library www.is.biu.ac.ilWebsite: Email: Ruthi.Tshop@mail.biu.ac.il   Tel. 972-3-5318163 .טל

Bar-Ilan University (RA)
Faculty of Humanities
Library of Information science

אוניברסיטת בר-אילן (ע"ר)
הפקולטה למדעי הרוח
הספריה ללימודי מידע

The test subjects' positions regarding the use of an automatic personal information classification software were split quite evenly. In average, there were 29% positive statements (standard deviation 36%), 35% hesitant statements (standard deviation 26%) and 37% negative statements (standard deviation 38%). According to the researcher's impression, 20% expressed a positive position and 35% expressed a pretty positive position.

Prominent categories of a positive nature were: 'Should there be an automatic classification software, I'd want to use it' (5 test subjects); 'Automatic classification of personal information answers an existing need' (5 test subjects). Prominent categories of a hesitant nature were: 'I'd use an automatic classification software while testing it, until I trust its abilities' (10 test subjects); 'I'd like to use an automatic classification software, provided it included certain features' (5 test subjects). A prominent category of a negative nature was: 'I prefer human decision-making to a computer software' (6 test subjects). Also, no unequivocal connection was found between the test subjects' positions regarding the use of a software for automatic classification of personal information and their user experience during the simulation experiment.

The test subjects mentioned several features for such a software, which they find important or that will give it added value. The main desired features were: that the software needs to be dynamic, that it should notice changes in the user's work patterns and that it should learn from mistakes (14 test subjects), as well as using a log and other control mechanisms.

The analysis of the questionnaire shows that most respondents see the value in a software for automatic classification of personal information (80%) and a similar number expressed their will to use such software (69%).

_____

Library of Information Science   הספריה ללימודי מידע
Bar-Ilan University, Ramat-Gan, Israel   אוניברסיטת בר-אילן, רמת גן, ישראל 52900
/library www.is.biu.ac.ilWebsite: Email: Ruthi.Tshop@mail.biu.ac.il   Tel. 972-3-5318163 .טל

Bar-Ilan University (RA)
Faculty of Humanities
Library of Information science

אוניברסיטת בר-אילן (ע"ר)
הפקולטה למדעי הרוח
הספריה ללימודי מידע

Accepting the results and conclusions should be done with reservations, since the experiment has limitations that affect the evaluation of the users' experience. A main limitation of the research was that the test subjects who agreed to participate in the experiment already had a positive or indifferent attitude towards automatic classification to begin with. People who fear automatic classification refused to take part in the experiment, and so the selection of test subjects is biased and it's hard to deduce from it on the entire population. Another limitation is the short duration of the experiment.

The research's findings indicate that the user experience of automatic classification of personal information, as expressed in the interviews, is derived from a variety of factors, of which success or failure of the classification is only one. Other factors were the manner in which the experiment was conducted and whether it induced positive feelings of trust and comfort or the other way around, as well as previous positions of the test subjects, that might have been positive or negative. The test subjects emphasized their need for control. Some related to an external locus of control and others to an internal locus of control. An important aspect that strengthens the sense of control was the trust the test subjects felt regarding the classification that was performed. Feelings of comfort or discomfort during the experiment were related to control or loss of control, and also to the invasion of personal space. Other negative feelings were related to information retrieval, and could be attributed to a lack of episodic memory of the file's classification by the test subjects.

The test subjects also referred to aspects of control, trust and comfort in the context of using a fictional software program for automatic classification of personal information. This matches previous research findings regarding the adoption of information technologies in general and

_____

Library of Information Science   הספריה ללימודי מידע
Bar-Ilan University, Ramat-Gan, Israel   אוניברסיטת בר-אילן, רמת גן, ישראל 52900
/library www.is.biu.ac.ilWebsite: Email: Ruthi.Tshop@mail.biu.ac.il   Tel. 972-3-5318163 .טל

Bar-Ilan University (RA)
Faculty of Humanities
Library of Information science

אוניברסיטת בר-אילן (ע"ר)
הפקולטה למדעי הרוח
הספריה ללימודי מידע

particularly the adoption of agents. Other reservations brought up by the test subjects regarded the software's feasibility.

When designing such software, several requirements should be taken into consideration in order to improve the sense of security, the feelings of control and comfort the user experiences, and most importantly: the software needs to be highly dynamic; to adjust its behavior according to the user's behavior and learn from the user if he overrides the decisions made by the agent; to give the user control over the agent's functions, while assigning more and more responsibilities to the agent, at a pace which is convenient to the user; and to provide clear and accessible feedback and control tools for monitoring the agent's work.

System No.: 002387809

_____

Library of Information Science   הספריה ללימודי מידע
Bar-Ilan University, Ramat-Gan, Israel   52900 אוניברסיטת בר-אילן, רמת גן, ישראל
/library www.is.biu.ac.ilWebsite: Email: Ruthi.Tshop@mail.biu.ac.il   Tel. 972-3-5318163 .טל