INTERACTING WITH INFORMATION CHALLENGES IN HUMAN–COMPUTER INTERACTION AND INFORMATION RETRIEVAL

Andreas Holzinger
Medical University Graz, Institute of Medical Informatics, Statistics and Documentation, Research Unit HCI
Graz, Austria

ABSTRACT

Today, we are confronted with a flood of information. Research in Human-Computer Interaction (HCI) and Information Retrieval (IR) has both long been working to develop methods that help people to identify, extract and understand useful information from data. The two fields, however, take very different perspectives in tackling the problem; and historically, they have had little collaboration. Let us for example look into the area of health: Medical professionals are faced with an increasing quantity of highly detailed, complex and non-standardized data at the press of a button, however, the time available to make decisions is the same as before the advent of such technological advances. According to (Gigerenzer, 2008) a typical medical doctor has approximately five minutes to make a decision. When everything turns out well no one complains; however, when something goes wrong, solicitors have nearly indefinite time to figure out whether and why a wrong decision has been made. The goal is to support medical professionals to interactively analyse information properties and visualize the most relevant parts without getting overwhelmed. The challenge is to bring HCI & IR to work together and hence reap the benefits, so that we can benefit medicine and health care even more.

KEYWORDS

Information Retrieval, Human-Computer Interaction, usability, complex data, complexity

1. INTRODUCTION AND MOTIVATION

Since the beginning of human existence, humankind has sought, organized and used information as it evolved patterns and practices of human information behaviour (Spink & Currier, 2006). Although the accessibility of information represents an important cultural advance, it also introduces a new challenge: retrieving only relevant information. To extract relevant information out of the vast complexity of data is the central quest of the modern information society.

However, accessing information is not always an easy task because we are dealing with the real world – where more data does not necessarily mean more information and more information is not always more knowledge. The challenge is that we have to consider the situation, the context. A further challenge is that mobile devices will be the primary tools in the future (Anderson & Rainie, 2008) and according to (Tsai et al., 2010) in mobile information retrieval there typically exist two main parts with their typical research fields: *Context Awareness* and *Content Adaption*.

Context Awareness deals with the fact that smart embedded devices have features to recognize the situation the user of the device is in at the moment. That means time, location, social status (social network).

Content Adaption mainly deals with how to present a user-friendly visualization of the results of information requests. Some examples on work for small screens can be found in (Noirhomme-Fraiture et al., 2005), (Sweeney & Crestani, 2004), (Jones, Buchanan & Thimbleby, 2002).

A further challenge is based on the fact that only a small percentage of data is structured – most of the data is semi-structured, weakly structured or even unstructured. A common misconception is to confuse structure with standardization. While the closely related fields of Information Retrieval and Knowledge Discovery have developed intelligent (semi)automatic processes and algorithms to extract useful knowledge from rapidly growing amounts of data, these methods fail when data are weakly structured and there is the danger of modeling artifacts. Consequently, there are a lot of relevant research issues on the intersection of HCI and IR to help (medical) professionals to identify and extract useful information from data.

2. HUMAN INFORMATION BEHAVIOUR

It is important to understand the four main terms that are essential when discussing human information behaviour (Spink & Saracevic, 1998), (Spink & Currier, 2006):

Information Behaviour is the totality of human behaviour in relation to sources and channels of information, including both active and passive information seeking, and information use. Thus, it includes face-to-face communication with others, as well as the passive reception of information as in, for example, watching TV advertisements, without any intention to act on the information given.

Information Seeking Behaviour is the purposive seeking for information as a consequence of a need to satisfy some goal. In the course of seeking, the individual may interact with manual information systems (such as a newspaper or a library), or with computer-based systems (such as the World Wide Web).

Information Searching Behaviour is the 'micro-level' of behaviour employed by the searcher in interacting with information systems of all kinds. It consists of all the interactions with the system, whether at the level of human computer interaction (for example, use of the mouse and clicks on links) or at the intellectual level (for example, adopting a Boolean search strategy or determining the criteria for deciding which of two books selected from adjacent places on a library shelf is most useful), which also involves mental acts, such as judging the relevance of data or information retrieved.

Information Use Behaviour consists of the physical and mental acts involved in incorporating the information found into the person's existing knowledge base. It may involve, therefore, physical acts such as marking sections in a text to note their importance or significance, as well as mental acts that involve, for example, comparison of new information with existing knowledge.

It is essential to understand the cognitive and perceptual abilities of the end users (Holzinger, Searle & Nischelwitzer, 2007). The best way is to understand the mental models of the respective end users. Mental models are defined as "cognitive representations of a problem [or information] situation or system" (Marchionini & Shneiderman, 1988), (Calero-Valdez et al., 2010). Some studies examined the role of users' mental model of an IR system in contributing towards search results and they argued that end users must have an appropriate mental model of a system in order to be able to use it to its full potential (Ahmed, McKnight & Oppenheim, 2004). Generally, the problem is that end users lack understanding of how the system operates. A good example is the study of (Dimitrioff, 1992), dealing with the relationship between the mental model of users and their search performance using a university search system. According to Dimitrioff, an accurate mental model of an online catalogue included eight components: the content of the database; the interactive nature of the system; the availability of more than one database; knowledge of multiple fields within records; knowledge of multiple indexes and/or inverted files; Boolean search capability; keyword search capability; and the use of a controlled vocabulary. The main barrier when studying mental models is the fact that they are not directly observable but must be studied by observing users' behaviour and is therefore hard to identify. Cognitive data such as users' knowledge, experience and expectations and how they cope with their information problem and interact with the system and interfaces are very important for the understanding of users' models of such systems.

3. USABILITY IN INFORMATION RETRIEVAL

The first IR systems in the early 1970s allowed searches via command interfaces. The major disadvantage of such interfaces is the fact that the users must be familiar with the command language of the system to use it effectively (Hawkins, 1981), giving skilled experts the advantage.

Despite all improvements in end user interfaces, recent studies reported that web-based interfaces are still difficult to use and the need for better IR interface designs is still remaining (Ahmed, McKnight & Oppenheim, 2009). This work also emphasizes individual differences when using search engines and identified following main influence keys:

Search Experience: Describes the fact that users who are used to search engines are using a broader variety of the query language they depend on. Having search engine experience clearly has a positive impact und the user's search performance.

Knowledge: Knowledge about the topic enables the user to use more synonyms and combination of search terms to fulfill the information requirements. Generally Search Experience has to be presumed before the Knowledge factor takes effect.

Academic Background: Users with an academic background in science and engineering have a better search performance than people from the humanties given the same level of expertise. Interestingly experienced female searchers had better results than experienced male searchers.

Users Age: Older users with the same computer experience as younger ones had a lower success rate. On the other hand, a certain level of language understanding has to be assumed so older elementary students had better results than younger ones. Searchers with computational experiences overcome those with less computational background. Obviously computer experience influences the use of search engines in a positive way. Summarizing, all these factors influence the users in the searching strategy and experience in a positive or negative way. So there is a need for search engines that also take account the *personal context of* the user, in order to optimize their experience and success in fulfilling the information requirements. Consequently, an ideal system should make use of the abilities of both the human and the computer in tasks to which they are best suited, and to provide explanations of the data for enabling insights (Beale, 2007).

4. COMPLEXITY AS MAIN CHALLENGE

Complexity is our main challenge, because most of our data is weakly structured or even unstructured and there is always the danger of modelling artefacts (which can then lead to wrong decisions). A good example are medical documents: The broad application of enterprise hospital information systems amasses large amounts of medical documents, which must be reviewed, observed and analyzed by human experts (Holzinger, Geierhofer & Errath, 2007). All essential documents of the patient records contain at least a certain portion of data which has been entered in non-standaridzed format (wrongly called *free-text*) and has long been in the focus of research (Gell, Oser & Schwarz, 1976). Although text can be *created* simple by the end-users, the support of automatic analysis is extremely difficult (Gregory, Mattison & Linde, 1995), (Holzinger et al., 2000), (Lovis, Baud & Planche, 2000). It is likely that some interesting and relevant relationships remain completely undiscovered, due to the fact that relevant data are scattered and no investigator has linked them together manually (Smalheiser & Swanson, 1998), (Holzinger et al., 2008).

Consequently, a major research area is on how to extract *knowledge* from this weakly or unstructured data. When we talk about structures, we will see some really interesting aspects of structures, in both microcosmos and macrocosmos.

A good example of a data intensive and highly complex microscopic structure is a yeast protein network. Yeasts are eukaryotic micro-organisms (fungi) with 1,500 currently known species, estimated to be only 1% of all yeast species. Yeasts are unicellular, typically measuring 4 µm in diameter. The first protein interaction network was published by (Jeong et al., 2001). The problem with such structures is that they are very big and that there are so many. A great challenge is to find unknown structures (structural homologies, see e.g. (Jornvall et al., 1981)) amongst the enormous set of uncharacterized data. Let us illustrate this process with a typical example from the life sciences: X-ray crystallography is a standard method to analyse the arrangement of objects (atoms, molecules) within a crystal structure. This data contains the mean positions of the entities within the substance, their chemical relationship, and various others and the data is stored in a Protein Data Base (PDB). This database contains vast amounts of data. If a medical professional looks at the data, he or she sees only lengthy tables of numbers.

However, by application of a special visualization method, such structures can be made graphically visible and the medical professionals can understand these data more easily and most of all they can gain knowledge – for instance, it may lead to the discovery of new, unknown structures in order to modify drugs, and consequently to contribute to enhancing human health. The transformation of such information into knowledge is vital for the prevention and treatment of diseases (Wiltgen & Holzinger, 2005), (Wiltgen, Holzinger & Tilz, 2007).

To demonstrate that not only natural processes have such structures there is a nice example from (Hurst, 2007) which shows a visualization of the blogosphere (cf. also with (Leskovec et al., 2007)): The larger, denser part of the blogosphere is characterized by socio-political discussion – the periphery contains some topical groupings. By showing only the links in the graph, we can get a far better look at the structure than if we include all the nodes.

A further example is from viral marketing. The idea is to spread indirect messages, which suggests spreading farther. If you press the *like*-button in Facebook – a similar process starts to an epidemic in medicine – a illness spreading through a population.

(Aral, 2011) calls it *behaviour contagion* and it is of much importance to know how behaviour can spread. We can mine masses of social network data in order to gain knowledge about the contagion of information. This is of particular interest for the health area.

5. CONCLUSION AND FUTURE OUTLOOK

Successful information retrieval systems will be those that bring the designer's model into harmony with the user's mental model. We can conclude that combining HCI with IR will provide benefits to our e-Society. Most of all, we must bridge Science and Engineering to answer fundamental questions on what is information and on (how) we can build such systems simply. Important future research aspects include:

1) research on the physics of information to contribute to fundamental research;

2) considering temporal and spatial information, in networks spatially distributed components raise fundamental issues on information exchange since available resources must be shared, allocated and re-used – Information is exchanged in both space AND time for decision making, therefore timeliness along with reliability and complexity constitute main issues and are most often ignored;

3) We still lack measures and meters to define and appraise the amount of information embodied in structure and organization – for example entropy of a structure;

4) considering information transfer: how can we assess, for example, the transfer of biological information;

5) Information and Knowledge: In many scientific contexts we are dealing only with data – without knowing precisely what these data represent. What is semantic information and how can we characterize it?

6) and most of all, we must gain value out of data – making data valuable.

Human-Computer Interaction and Information Retrieval (HCI & IR) is dedicated to contribute towards these challenges in their own ways; the challenge is to get them to do it collaboratively, and hence benefit medicine and health care even more.

ACKNOWLEDGEMENTS

I am grateful for inspirations and help from Marcus Bloice, Sue Bogner, Philippe Palanque, Vimla Patel, Gig Searle, Klaus-Martin Simonic, Harold Thimbleby and Martina Ziefle.

REFERENCES

Ahmed, S. M. Z., McKnight, C. & Oppenheim, C. (2004) A study of users' performance and satisfaction with the Web of Science IR interface. *Journal of Information Science,* 30, 5, 459-68.

Ahmed, S. M. Z., McKnight, C. & Oppenheim, C. (2009) A review of research on human-computer interfaces for online information retrieval systems. *Electronic Library,* 27, 1, 96-116.

Anderson, J. Q. & Rainie, L. (2008), The Future of the Internet III. Online available: http://www.pewinternet.org/Reports/2008/The-Future-of-the-Internet-III.aspx, last access: 19.05.2011

Aral, S. (2011) Identifying Social Influence: A Comment on Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science,* 30, 2, 217-223.

Beale, R. (2007) Supporting serendipity: Using ambient intelligence to augment user exploration for data mining and Web browsing. *International Journal of Human-Computer Studies,* 65, 5, 421-433.

Calero-Valdez, A., Ziefle, M., Alagöz, F. & Holzinger, A. (2010) Mental Models of Menu Structures in Diabetes Assistants. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (Ed.) *Lecture Notes in Computer Science (LNCS 6180).* Berlin, Heidelberg, New York, Springer, 584-591.

Dimitrioff, A. (1992) Mental models theory and search outcome in a bibliographic retrieval system. *Library and Information Science Research,* 14, 2, 141-155.

Gell, G., Oser, W. & Schwarz, G. (1976) Experiences with the AURA Free Text System. *Radiology,* 119, 105-109.

Gigerenzer, G. (2008) *Gut Feelings: Short Cuts to Better Decision Making* London, Penguin.

Gregory, J., Mattison, J. E. & Linde, C. (1995) Naming Notes - Transitions from Free-Text to Structured Entry. *Methods of Information in Medicine,* 34, 1-2, 57-67.

Hawkins, D. T. (1981) Online information retrieval systems *Annual Review of Information Science and Technology,* 16, 171-208.

Holzinger, A., Geierhofer, R. & Errath, M. (2007) Semantic Information in Medical Information Systems - from Data and Information to Knowledge: Facing Information Overload. In: Tochtermann, K., Haas, W., Kappe, F. & Scharl, A. (Eds.) *Procedings of I-MEDIA '07 and I-SEMANTICS '07.* Graz, 323-330.

Holzinger, A., Geierhofer, R., Modritscher, F. & Tatzl, R. (2008) Semantic Information in Medical Information Systems: Utilization of Text Mining Techniques to Analyze Medical Diagnoses. *Journal of Universal Computer Science,* 14, 22, 3781-3795.

Holzinger, A., Kainz, A., Gell, G., Brunold, M. & Maurer, H. (2000). *Interactive Computer Assisted Formulation of Retrieval Requests for a Medical Information System using an Intelligent Tutoring System*. World Conference on Educational Multimedia, Hypermedia and Telecommunications ED-MEDIA 2000, Montreal, Charlottesville (VA): AACE, 431-436.

Holzinger, A., Searle, G. & Nischelwitzer, A. (2007) On some Aspects of Improving Mobile Applications for the Elderly. In: Stephanidis, C. (Ed.) *Coping with Diversity in Universal Access, Research and Development Methods in Universal Access, Lecture Notes in Computer Science (LNCS 4554).* Berlin, Heidelberg, New York, Springer, 923-932.

Hurst, A. (2007), Data Mining: Text Mining, Visualization and Social Media. Online available: http://datamining.typepad.com/data_mining/2007/01/the_blogosphere.html, last access: 2011-05-10

Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. (2001) Lethality and centrality in protein networks. *Nature,* 411, 6833, 41-42.

Jones, M., Buchanan, G. & Thimbleby, H. (2002) Sorting out searching on small screen devices. *Lecture Notes in Computer Science LNCS 2411*. Berlin, Heidelberg, New York, Springer, 81-94.

Jornvall, H., Carlstrom, A., Pettersson, T., Jacobsson, B., Persson, M. & Mutt, V. (1981) Structural Homologies between Pre-albumin, gastrointestinal Pro-Hormones and other Proteins. *Nature,* 291, 5812, 261-263.

Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N. & Hurst, M. (2007) *Patterns of Cascading Behavior in Large Blog Graphs.* Philadelphia, Siam.

Lovis, C., Baud, R. H. & Planche, P. (2000) Power of expression in the electronic patient record: structured data or narrative text? *International Journal of Medical Informatics,* 58, 101-110.

Marchionini, G. & Shneiderman, B. (1988) Finding Facts versus Browsing Knowledge in Hypertext Systems. *IEEE Computer,* 21, 1, 70-80.

Noirhomme-Fraiture, M., Randolet, F., Chittaro, L. & Custinne, G. (2005). *Data Visualizations on small and very small screens*. Proceedings of ASMDA 2005: 11th International Symposium on Applied Stochastic Models and Data Analysis, ENST, Brest (France).

Smalheiser, N. R. & Swanson, D. R. (1998) Using ARROWSMITH: a computer-assisted approach to formulating and assessing scientific hypotheses. *Computer Methods and Programs in Biomedicine,* 57, 3, 149-153.

Spink, A. & Currier, J. (2006) Towards an evolutionary perspective for human information behavior - An exploratory study. *Journal of Documentation,* 62, 2, 171-193.

Spink, A. & Saracevic, T. (1998) Human-computer interaction in information retrieval: nature and manifestations of feedback. *Interacting with Computers,* 10, 3, 249-267.

Sweeney, S. & Crestani, F. (2004) Supporting searching on small screen devices using summarisation. *Lecture notes in computer science LNCS 2954, Mobile HCI, Udine*. 187-201.

Tsai, F. S., Etoh, M., Xie, X., Lee, W. C. & Yang, Q. (2010) Introduction to Mobile Information Retrieval. *Ieee Intelligent Systems,* 25, 1, 11-15.

Wiltgen, M. & Holzinger, A. (2005) Visualization in Bioinformatics: Protein Structures with Physicochemical and Biological Annotations. In: Zara, J. & Sloup, J. (Eds.) *Central European Multimedia and Virtual Reality Conference (availbale in EG Eurographics Library).* Prague, Czech Technical University (CTU), 69-74.

Wiltgen, M., Holzinger, A. & Tilz, G. P. (2007) Interactive Analysis and Visualization of Macromolecular Interfaces Between Proteins. In: Holzinger, A. (Ed.) *HCI and Usability for Medicine and Health Care. Third Symposium of the Workgroup Human-Computer Interaction and Usability Engineering (HCI&UE) of the Austrian Computer Society, USAB 2007, Lecture Notes in Computer Science (LNCS 4799).* Berlin, Heidelberg, New York, Springer, 199-212.