USERS' IMAGE SEEKING BEHAVIOUR IN MULTILINGUAL ENVIRONMENTS: A GROUNDED THEORETICAL APPROACH

EVGENIA VASILAKAKI

A thesis submitted in partial fulfilment of the requirements of the Manchester Metropolitan University for the degree of Doctor of Philosophy

Department of Information and Communications
the Manchester Metropolitan University
May 2011

Acknowledgements

Abstract

This Thesis aims to explore users' image seeking behaviour in online multilingual environments. In particular, it focuses on identifying users' actions/interactions- and on enquiring into users' rationales and justifications for their actions. In addition, it aims to identify the factors which have affected and/or informed users' image seeking behaviour in multilingual environments. In this context, an inductive research approach and specifically 'Grounded Theory' as a methodology were adopted. A mixture of four different methods (questionnaire, observation, retrospective thinking aloud and interview), both qualitative and quantitative, were employed for data collection.

For the purposes of this thesis, two studies were conducted. In particular, an Exploratory Study served as the means for gaining a first insight into users' image seeking behaviour. This in turn, informed the design and conduction of the Main Study. A procedural analysis of the data collected was adopted focusing on identifying users' actions/interactions, their rationales behind these actions and finally the consequences of these actions.

A substantive theory of users' non-linear image seeking behaviour in a multilingual context emerged from the data. In particular, twenty seven concepts and numerous codes emerged and were accordingly assigned to the three distinct areas: conditions, actions and consequences. In the quest for a higher conceptual analysis, four conceptual categories were identified in the concepts: *Knowledge of Languages*, *Query Domain*, *System* and *Search*. Diagrams were used to illustrate the theory found in data, to demonstrate the relationships among concepts and to gain analytical distance from data.

This thesis contributes to our understanding of the diversity and complexity of users' searching experiences in multilingual environments and fills a gap in the relevant literature by providing substantive insights into the relationship between the interface and users' actions. In addition, it contributes to learning how users' perceive, adopt and adjust MLIR systems to their needs and their search behaviour. This in turn, contributes to informing the design of efficient MLIR systems which support effective search behaviour and performance.

Contents

List of Figures

List of Tables

List of Abbreviations

CLEF Cross Language Evaluation Forum

CLIA Cross Language Information Access

CLIR Cross Language Information Retrieval

ExS   Exploratory Study

FIRE Forum for Information Retrieval Evaluation

GGT   Glaserian Grounded Theory

GT   Grounded Theory

HCI   Human Computer Interaction

iCLEF interactive CLEF

IR   Information Retrieval

IS   Information Systems

ISB   Information Seeking Behaviour

MLIA Multilingual Information Access

MLIR Multilingual Information Retrieval

MlM   Multilingual Mode

MMR   Mixed Methods Research

MMU   Manchester Metropolitan University

MnM   Monolingual Mode

MnS   Main Study

NTCIR NII Test Collection for IR Systems

RTA   Retrospective Thinking Aloud

SGT   Straussian Grounded Theory

TIA   Translingual Information Access

TIR   Translingual Information Retrieval

TREC   Text REtrieval Conference

WWW   World Wide Web

Chapter 1

Introduction

The wide spread adoption of the World Wide Web (WWW), the passage to the Semantic Web, the increasing amount of information produced daily in various languages and the rapid growth of non-English native speaking users demonstrate a need for developing systems for supporting Multi-Lingual Information Retrieval (MLIR) and Access (MLIA). A variety of different techniques for implementing and organizing different linguistic resources and achieving improved retrieval results have been tested and evaluated. These efforts aimed to provide the best possible system for the average web user to search for and retrieve the desired information regardless of the language in which it was written.

In this context, research was focused initially on developing MLIR systems with little regard for the user. That is development was largely technology driven. More recently, the need to recognize the user's point of view in the gathering of requirements and the evaluation of systems has been understood. But even here, such developments took place in a limited way. Specifically, user participation was limited to involvement in the evaluation of relevance of the retrieveditems and therefore enabling the calculation of relevance and recall rather than a more open- ended concern for what users might actually be doing with such systems. In turn, the small number of studies which employed users with the view to investigating their behaviour adopted mainly quantitative methods. This, in turn, was based on the view that defining specific tasks for users, and approaching the study of their behaviour in a quantitative way (and aggregating the results) will provide with 'clean' results which explain how an interface functions (Karat and Karat, 2003).

Research into image retrieval in multilingual environments was a subset of this research agenda and as such shared the same assumptions and, it is suggested, exhibited the same problems. To reiterate, this consisted of predetermined assumptions about the 'problem'- defined in terms of accuracy and speed of retrieval- and the adoption of quantitative methods to produce aggregate results. None of this reflected developments

elsewhere, which entailed a 'turn to the social' and typically involved, as we shall see, a more qualitative approach.

Existing studies both in information and image retrieval in multilingual environments, it is argued, failed to capture certain problems that arise- problems, which do become visible under different methodological auspices. In particular, elements such as the complexity of users' information seeking behaviour and the relevance of factors affecting it may have been initially underestimated or undervalued. In addition, variations in the nature of the user were considered less important than features of the interface itself. Further, the dominance of the hypothetico- deductive model has resulted in largely statistical results which produce aggregate responses and hence tell us little or nothing about the way in which we might discriminate different user behaviours and the reasons for them.

In contrast, the adoption of an inductive approach and the generation of theory from data using a mixture of both qualitative and quantitative methods- predicated on the 'grounded theoretical' approach became- the methodological basis for the studies reported in this thesis. The research described below aimed, by adopting a 'mixed method' approach utilizing both qualitative and quantitative results, to fill in this gap by focusing on users' image seeking behaviour in multilingual environments from a user perspective.

1.1 Aim

This study aimed to investigate users' image seeking behaviour in multilingual environments. In particular, it focused on identifying users' actions/interactions- and on enquiry into users' rationales and justifications for their actions. In addition, it aimed to identify the factors which have affected and/or informed users' image seeking behaviour in multilingual environments.

The study is not about testing or verifying a hypothesis since there was no relevant pre-existing literature in users' image seeking behaviour in multilingual environments from which to work. It is predicated on the view that valid evaluation of systems must entail the study of 'systems in use'. In other words, it is based on a socio-technical view of the problems relating to the use of such systems. Interfaces cannot be properly evaluated independently of the way in which people actually use them.

1.2 Contribution to Knowledge

This thesis contributes to our understanding of users' searching experiences in multilingual environments and fills a gap in the relevant literature by providing substantive

insights into the relationship between interface and action. In particular, it contributed to providing an insight into users' image seeking behaviour focusing on users' rationales and justifications for their actions. In particular, the procedural analysis, presentation of findings and of the emergent theory provided a valuable understanding of the search process which users employ while searching. This in turn, contributed to learning how users' perceive, interpret and employ MLIR systems and adjust systems to their needs and their general search characteristics and not the other way around. This understanding informed the design of efficient MLIR systems which support effective search behaviour and performance without placing additional or unnecessary cognitive demands on users wide acceptance of these systems.

1.3   Outline of Chapters

This thesis consists of six chapters. An outline of these chapters is presented here.

In Chapter 2, a brief review of the Information Seeking Behaviour (ISB) literature is presented, followed by a more focused review of the literature on Multilingual Information Retrieval (MLIR). Related terms are defined followed by a brief history of the technological developments achieved to create systems which supported MLIR. In turn, the focus is on the contributions made by the existing literature. It will be suggested that, in the main, this literature has been either system- centred, with no role for user studies, or user- centred in a limited way, the limits being defined by hypothetico-deductive, experimental approaches which employ mainly quantitative methods.

The review of literature concludes with a review of the literature on image retrieval in multilingual environments which, is the research area of this study. It will be suggested that issues of 'trust' and 'confidence', increasingly raised as issues in user behaviour in other literature (e.g. social sciences see (Deutsch, 1958), psychology see (Erikson, 1968)), may be relevant here. Indeed, it became apparent during the course of the studies that trust and confidence were important factors in understanding user behaviour. Finally, a brief outline of Human Computer Interaction research and how it evolved is presented because it also informed this study. In particular, it provided the grounds for the move to a more qualitative approach and for an approach which allowed some redefinition of the research problem as the study proceeded. The appreciation of 'context' found in the HCI literature was a basis for understanding that the complexity of user behaviours cannot be easily grasped without an understanding of contextual factors. The focus of studies turned, to coin a phrase, from 'user' to 'using'.

In Chapter 3, the contrast between existing hypothetico- deductive research approaches, and a more inductive and 'grounded' approach is examined and the latter justified. In addition, Grounded Theory (GT) is illustrated and the choice of the Straus-

sian GT is explained as the predefined methodology for enabling a successful collection and analysis of data. An outline of the research setting is then presented which in turn determined the kind, number and way methods were implemented. Mixed Methods Research (MMR) is then described. To be clear, where GT is a methodology- an approach to analysis and the generation of theory- it does not prescribe any particular set of methods. Indeed, it is argued that it is suitable for use with any combination of research methods (Strauss and Corbin, 1998). The generation of theory can be based on any number of data collection methods and processes of 'triangulation' may be useful in testing the validity of observations. MMR is precisely useful for these purposes. Moreover, the specific methods of questionnaire, observation, retrospective thinking aloud and interview and the way they were implemented in this work are briefly illustrated. Furthermore, the aim and specific elements of the Exploratory Study (ExS) are explained. In particular, details regarding the test object, user sampling, task, methods designed and employed, data collection and analysis process and ExS' limitations are presented. Finally, the elements of Main Study (MnS) are presented in much the same way.

In chapter 4, the findings from the ExS are presented. In particular, the phenomenon of users' searching experiences in multilingual environments was identified. In this context, the concepts of users' usage of modes, users thinking about languages and paying attention to translations were identified and described. A brief discussion of findings in relation to the limited relevant literature is undertaken. As a result, the role of ExS was three fold; a) ExS provided the setting for testing the design and implementation of the methods employed; b) ExS provided an insight into users' image seeking behaviour; and c) its findings formulated the basis for the emergence of substantive theory. In particular, it provided some hints of factors such as trust and confidence in searching and language skills which worth exploring of how and if they have affected users behaviour.

In Chapter 5, MnS' Findings are presented in a procedural way. In particular, a description of the criteria used for identifying the cases (in keeping with a GT approach) of users' seeking behaviour in a procedural way is illustrated. Then users were assigned to each case and one user's image seeking behaviour in FlickLing for each case is presented in detail. As such, diversity of users' image seeking behaviour in FlickLing was allowed to emerge and complexity among users' behaviour were shown. The identification of the cases and the procedural presentation of analysis enabled the generation of substantive theory.

In Chapter 6, the way Substantive Theory emerged is presented. In particular, each of the emerged concepts and their codes are illustrated and discussed with the relevant literature where applicable. The way the four conceptual categories of Knowledge

of Languages, Query, Search and System emerged is described and diagrams are employed to demonstrate the relationships found among concepts. Finally, the emerged substantive theory of users' non-linear image seeking behaviour in multilingual context is discussed with the relevant literature of models in Information Seeking Behaviour research area. In particular, Foster (2004, 2005, 2006)'s model was used in an attempt to explain some of the complexity and the category relations emerging from the research in this thesis.

In the final Chapter 7, some concluding remarks are presented regarding the research area of the study. In particular, the way in which the adopted research approach, methodology and methods led to the exploration of users' image seeking behaviour in multilingual environments is illustrated. The way in which the conclusions drawn from Exploratory Study which subsequently informed Main Study and the contribution of procedural analysis in identifying the Paradigm and the conceptual categories and conclusions drawn from the emerged substantive theory are all explored. Finally, ways of taking this research forward are proposed.

Bibliography

Deutsch, M. December (1958). Trust and suspicion. *The Journal of Conflict Resolution*, 2(4):265–279.

Erikson, E. (1968). *Identity, youth and crisis.* New York, US: Norton.

Foster, A. (2004). A nonlinear model of information seeking behaviour. *JASIST*, 55: 228–237.

Foster, A. (2005). Nonlinear information seeking. In Fisher, K., Erdelez, S., and MacKechnie, L..(editors), *Theories of information behavior*, pages 254–258. ASIST monograph series.

Foster, A. (2006). A non-linear perspective on information seeking. In Spink, A. and Cole, C..(editors), *New directions in human information behavior*, pages 155–170. Springer.

Karat, J. and Karat, C. (2003). The evolution of user-centered focus in the human-computer interaction field. *IBM Syst. J.*, 42(4):532–541.

Strauss, A. and Corbin, J. September (1998). *Basics of Qualitative Research : Techniques and Procedures for Developing Grounded Theory.* SAGE Publications.