

Users' Image Seeking Behaviour in Multilingual Environments: experience in combining qualitative and quantitative data

Evgenia Vassilakaki¹, R.J. Hartley¹,
Frances Johnson¹ and Dave Randall²

1 evgenia.vassilakaki@student.mmu.ac.uk

r.j.hartley@mmu.ac.uk

f.johson@mmu.ac.uk

Dept. Information & Communications

Manchester Metropolitan University

United Kingdom

2 d.randall@mmu.ac.uk

Dept. Sociology

Manchester Metropolitan University

United Kingdom

Abstract: The volume of information available on the Web in various languages and the growing number of the non-English native speaking users generate a need for developing and providing Cross Language Information Retrieval (CLIR) systems. The efficient development of such systems depends greatly on gaining a better insight into users' interactions and search behaviours in multilingual environments. In this context, the decision about which method/s to employ is considered essential. This study reports on the reasons why a mixed methods research was adopted; the decision for implementing questionnaire, observation, retrospective thinking aloud and interviews; the way in which they were combined and the values and limitations posed by such an implementation.

Keywords: Multilingual Information Retrieval, Mixed Methods Research, Qualitative/ Quantitative Methods, CLIR Systems, User Image Seeking Behaviour

1. Introduction

The increasing amount of information accessible on the Web in various languages and the rapid growth of non-English native speaking users demonstrate a need for developing systems in Organizations, Libraries, Archives for supporting Cross Language Information Retrieval (CLIR) and Access (CLIA). A variety of different techniques for implementing and organizing different linguistic resources and achieving better retrieved results have been tested and cross-evaluated. All these efforts aimed to provide the best possible system for the average web user to search and retrieve the desired information regardless of the language written. In this process, little attention has been given to user interaction and behaviour with these systems which are considered to be essential if we are to develop CLIR systems that support effective search behaviour and performance.

In this context, the decision on the specific methods to be used and the way in which they are to be implemented is considered important to obtain insights into users' search behaviours when developing a CLIR system. In particular, this study intends to report on the specific four methods employed, the reasons why these have been chosen, the way they were implemented and the value and limitations that these posed when developing CLIR systems that support effective search behaviour and performance.

The remainder of the paper is structured as follows: a brief description of the way that the study was set and carried out is given in section 2. We report on the reasons why a mixed methods research was adopted in this study in section 3. A description of the way the four different methods were implemented and the strengths and weaknesses of each are reported in sections 4, 5, 6 and 7

consequently. A discussion on the strengths and weaknesses in regards of the generated findings is followed in section 8. Finally, we conclude summarizing the way that these methods could be used from a managerial perspective of decision making to provide efficient CLIR systems and services to users in section 9.

2. Study's Description

This study aims to draw insights on the value and limitations posed by the employment of both quantitative and qualitative methods when exploring users' information seeking behaviour in multilingual environments. As this study focuses on methods, the description of the way this study was executed will be kept brief. The full length on the findings on users' image seeking behaviour reported can be found in Vassilakaki et al. (2008).

This study adopted the FlickLing Interface, a multilingual search interface for the well-known web application Flickr as the test object. FlickLing was developed and offered by the organizers of the CLEF2008¹ to enable user behaviour experiments for the interactive track (iCLEF2008) to be performed Clough et al. (2008). FlickLing uses the Flickr collection to provide a realistic multilingual search setting for carrying out cross-language user behaviour experiments. Additional information can be found in Clough et al. (2008) and Peinado et al. (2008).

The task assigned to the users was to search across languages for the given three known, non-annotated images in FlickLing after login. The users did not know in advance, in which of the six languages supported by FlickLing (English, German, Dutch, Spanish, French, Italian) the images were described. As a result, users were forced to use all the features and functionalities of FlickLing to search for the images across languages.

Finally, a sample of ten (10) users, three male and seven female, aged 20 to 40, was used for the purposes of this study. All ten users were in some way or another related to Manchester Metropolitan University (MMU). Users were all coming from different cultural backgrounds, speaking a variety of native languages and having a different level of comprehension for a variety of foreign languages.

Each user on arrival was given general instructions regarding the study and the assigned task. They were then asked to fill in a questionnaire enquiring on personal details and prior experience and engage on completing the task. A system recording software was used to capture users' search behaviours while performing the task. The video produced was then played back to users asking them to describe their actions focusing on why they were behaving in this way and what they were thinking in retrospect. At the end, a brief, semi-structured interview was run with each user. All sessions were run individually, in the same lab and lasted approximately two hours each.

3. Mixed Methods Research

The use of different methods for collecting data and answering research questions arose due to a strong debate concerning the adoption and implementation of only qualitative or only quantitative methods. The idea of implementing methods from both categories has been initially faced with doubt and controversy in various disciplines. In the social and behavioural sciences, a number of reasons have been identified regarding the implementation of mixed methods in carrying out research with the same quality and validity as with the mono-method research. In particular, three areas can be identified in which mixed methods appear to be superior to single method approaches (Tashakkori and Teddi (2003): a) mixed methods research can answer research questions that the other methodologies cannot,

1 Cross Language Evaluation Forum (CLEF) <http://www.clef-campaign.org/>

b) mixed methods research provides better (stronger) inferences and c) mixed methods provide the opportunity for presenting a greater diversity of divergent views.

In this context, a variety of ways of combining qualitative and quantitative methods to address research questions have been defined. We are taking the view of Bryman (2008a) and (2008b), who identified the following ways based on a content analysis of empirical articles in the social sciences using mixed methods research:

- Triangulation or greater validity: is a way in which quantitative and qualitative methods are combined to triangulate findings in order to be mutually corroborated. The idea of triangulation was initially evolved from the work of Campbell and Fiske (1959) and later expanded by Denzin (1978) introducing four different types of triangulation methods, including data triangulation, investigator triangulation, theory triangulation and methodological triangulation.
- Offset: this way suggests that both quantitative and qualitative methods adopted to address the research questions have their own strengths and weaknesses, so the researcher by combining them can offset the weaknesses and draw on their strengths.
- Completeness: in this way the researcher uses both qualitative and quantitative methods to form a more comprehensive account of the research area.
- Process: qualitative methods provide a sense of process lacking from quantitative methods which focus on the structures in social life.
- Different research questions: it is based on the argument that different methods (quantitative/qualitative) address different questions as expressed by the authors themselves.
- Explanation: when one of the two methods is used to explain the other method's findings.
- Unexpected results: it is suggested that a combination of quantitative and qualitative methods could help to explain unexpected results generated by one method by employing the other.
- Instrument development: when qualitative methods are employed in order to gather necessary data for formulating the final questionnaire and scale items in order to achieve better wording of both questions and answers provided.
- Sampling: when one method is used to enable the sampling of respondents of cases.
- Credibility: it is based on the suggestion that employment of both methods (quantitative and qualitative) could enhance the integrity of findings.
- Context: refers to situations where qualitative methods are used to provide contextual understanding enabling at the same time the generalisation or identification of relationships among variables.
- Illustration: the use of qualitative methods to illustrate quantitative findings.
- Utility or improving the usefulness of findings: the combination of both methods could be more useful for practitioners and others.
- Confirm and discover: the use of qualitative data to form the research questions and quantitative methods to test their applicability in a specific case.
- Diversity of views: the combination of researchers and participants perspectives by employing quantitative and qualitative methods and identifying variables through quantitative method and meanings through qualitative.
- Enhancement or building upon quantitative/qualitative findings: the use of quantitative or qualitative methods to build on already obtained quantitative or qualitative data.
- Other/ unclear: no clear way of using mixed methods research was identified in some of the empirical articles in the social sciences after conducting a content analysis.
- Not stated: no specific or stated way of using mixed methods research was identified in some of the empirical articles in the social sciences after conducting a content analysis.

The decision to adopt and apply a mixture of quantitative and qualitative methods in this study falls in the offset and completeness categories. In particular, four different methods both qualitative and quantitative have been employed: a) questionnaires, b) observation, c) retrospective thinking aloud and d) interviews. These methods have been chosen for the following reasons: a) they extract different data (quantitative/ qualitative), b) they overlap ensuring that no relevant data will be lost and c) they enable the collection of both concrete and difficult to misinterpret data.

Apart from the mixture of the different methods employed, the order of these also proved significant in both collecting and verifying the data. In particular, the questionnaire and the observation sheet were providing the facilitators with hints on users' information seeking behaviour while retrospective thinking aloud and the interviews were clarifying the reasons why users were behaving in this way. In addition, the stress put on the users was kept in balance throughout the study minimising the possibilities of users feeling tired early or uncomfortable in the study.

An extended description of the way the four methods have been applied will follow, focusing on: a) the way they have been designed, b) the main reason for having been designed in this way and c) how these methods contributed to the collection and final analysis of the data. Moreover, for each method a series of strengths and weaknesses will be described based on the way these were applied and on the data collected and analysed.

4. Questionnaire

This study applied a questionnaire mainly for two reasons: a) to gather information about the user sample and b) to enable the formulation of questions to be asked during interviews regarding users' search behaviour and habits adopted in the specific study. It was employed more as a supplement to the main method of the study (retrospective thinking aloud) for collecting data and contributing to their further analysis.

In this context, the questionnaire was designed to extract data on the following distinct areas of the user sample: a) users characteristics, b) users' previous experience in searching on the web for images across languages and c) users' familiarity with Flickr web application. In particular, it contained mainly a set of closed questions and a few open questions when it was needed to justify further users' answers.

Users were asked at the beginning to specify both their native language and their level of comprehension for each of the six languages (English, German, Dutch, Spanish, French, Italian) supported in FlickLing and/or state their knowledge for any additional languages. These were important to establish if they are monolingual, bilingual or multilingual and their level of comprehension for each of these languages for reasons of competency in searching in these languages. Moreover, it enquired after users' experience in: a) searching images on the web and b) searching for images across languages. These enabled us to form an idea of their level of experience and focus on the reasons why they search for images in other languages.

In regards to prior usage of Flickr application, our interest focused on knowing how familiar users were with Flickr. It was thought that experience in Flickr would be a variable affecting users' image seeking behaviour since FlickLing was based on and using Flickr's image database.

Finally, three questions were included regarding users' demographic details such as age, gender, education to further identify the characteristics of the user sample of this study. In addition, these were placed at the end so that users felt more comfortable answering them rather than thinking why their personal-demographic details were requested at the beginning of the study.

This questionnaire was also used to form questions during interviews. In particular, the facilitator would go through the questionnaire when the user would have answered it and if any of the user's answers were contradictory to his search behaviour then the facilitator would ask why was that so. For example, when a user with a basic knowledge in French as stated in the questionnaire would receive a

French annotated image and would not search in French or use the provided translations, the facilitator would ask to find the reason of this behaviour.

The employment of the questionnaire enabled the definition of the user sample characteristics and users' previous experience in each of the sections of interest. This questionnaire was kept as brief, concrete and easy to fill in as possible, mainly: a) to minimize the hassle to which users had to go through in completing it, b) to relieve the overload from carrying out the whole experiment (employment of four different methods) and c) to constrain the time spent on full-filing the various elements of the study. At the same time, it enabled the credibility of the data gathered.

On the other hand, users were left alone to complete the questionnaire and as a result some of them did not answer all questions, withhold some information or provided contradictory data because they had either misunderstood, misinterpreted the questions or did not pay attention and filled it in very quickly.

5. Observation

The method of observation was also employed to form an overview of the users' information seeking behaviour while performing the given task. In this context, the facilitator would be able to follow up users' discussion and explanations during the retrospective thinking aloud process and could form the necessary questions during the interviews based on the observations made.

A specific form was designed to assist the work of the facilitator in forming specific questions during interviews regarding pre-selected research areas of the test object and the study. This was done in an attempt to shed light on specific users' behaviours on specific occasions. This form was categorized according to the areas that were to be examined. These were defined in accordance with the main aim of the study. Every category had a set of predefined questions/ remarks that the facilitator had to complete according to the user's behaviour each time and write additional comments. A set of three forms was assigned to each user, one for each image. These forms were coded according to the number assigned to each user and the order of the images (e.g. 01/01, 01/02, 01/03). These were to be used for forming the questions during individual interviews in order: a) to verify users' saying on retrospective thinking aloud, b) to shed more light in the provided answers and the reasons stated and c) to bring users' thoughts on the surface.

The observation sheet designed proved to be a very valuable tool for the facilitator to extract data and formulate questions on users' behaviours and actions during interviews. In particular, it helped the facilitator to focus on specific areas, write comments, remarks and relevant questions for each user. In addition, the facilitator would note down users' expressions while searching for the images that the facilitator would use during interviews to engage users in conversation asking him/her to explain why the user made that comment. On the other hand, the facilitator could miss the overall user interaction because he/she was focusing on specific elements of users' behaviour. In addition, the quantitative data gathered though concrete would have no meaning or could easily be misinterpreted without users' further comments and justifications of these actions. Finally, the users may have altered their search behaviour because they were observed, as a result this may have affected the data gathered though all possible precautions have been taken to avoid it.

6. Retrospective Thinking Aloud

Retrospective thinking aloud belongs to Think Aloud Protocols and it is a widely used method for usability testing of software and interfaces Ericsson, K.A., and Simon, H. (1985) and Haak, M., de Jong, M., and Schellens, P. J. (2003). This study employed retrospective thinking aloud aiming to

derive findings entirely on users' thoughts, comments and search behaviours rather than depending on data gathered through only quantitative methods and facilitator's observations, interpretations and questions. This method required that users would complete the given task with the testing object and describe their thoughts and actions afterwards which were screen captured using camtasia.

This method comes with many benefits for both the users and the facilitators. The users were left alone to work on the given task in their own way and at their own pace, with no time constraints and interact with the test object. This practice proved beneficial for the non-English native speakers since they could search thinking in their own language as they do in their daily lives. Most importantly, afterwards during retrospective thinking aloud they could elaborate more and explain better their actions because they had the necessary time and they were relieved from the stress of completing the task assigned. In addition, the users had the opportunity and time to feel comfortable with the environment, test object and the facilitator. In this way, the focus was on peoples' cognitive processes after having completed a specific task.

In addition, retrospective thinking aloud enabled the facilitator to gain insights on users' information seeking behaviours, identifying the reasons why users were behaving in this way and asking relevant questions during interviews. These were used for verifying what the users stated on their own or to enquire more details and justifications. In addition, the users rather than the facilitator pointed out the problems concerning the test object.

The use of a screen capture software in system recording users actions and playing it back to them in real time also contributed to the study. In particular, the users could easily reflect back to their thoughts and explain their actions leaving no room for misinterpretation. In this context, the users were providing information on their general search behaviour in an attempt to justify their actions.

On the other hand, the users were feeling tired during the retrospective thinking aloud because as a process it was time consuming since it was taking as much time as the users have already spent on fulfilling the task. As a result, they were expressing unease and tiredness with describing everything in detail because of the time already spent on the task. In addition, the majority of the users were trying to justify as clearly as possible why they were failing to find the image because the recording of their performance was making them feel uncomfortable. The users were thinking that they did not employ the necessary search mechanisms based on what they were seeing on the screen. They were even trying to find ways of how they could have search differently and how this could have affected the final outcome.

7. Interviews

The last part of the study consisted of small scale individual interviews with every user after the completion of the retrospective thinking aloud. The main goal of these interviews was to clarify: a) the answers provided in the questionnaires, b) the specific actions of the user's image seeking behaviour during the search session and c) the users' expressions during retrospective thinking aloud.

The interviews carried out were semi-structured and the questions asked derived from the questionnaire, the observation sheets and the retrospective thinking aloud process. As a consequence, the questions asked to the users varied but they were all focused on users' image seeking behaviour when searching across languages.

The interviews enabled the verification of specific sayings of the users during retrospective thinking aloud and of the users' answers in the questionnaire. In addition, the data gathered through observation were able to be interpreted as users were asked to provide explanations on specific actions noted on the observation sheet. On the other hand, the interviews were kept brief because users were already tired and they were not willing to spend a long time on answering additional questions.

8. Discussion.

The development of effective CLIR systems depends to a great extent on exploring and recording users' information seeking behaviour and the reasons which trigger these behaviours. The identification and employment of the most suitable methods to monitor these behaviours are essential in generating the necessary findings. This study employed a mixed methods research drawing from the view of Tashakkori and Teddlie (2003) that mixed methods enable better (stronger) inferences and can answer questions that single methods cannot.

In this context, a combination of four different methods, both quantitative and qualitative have been adopted to generate findings. These methods were employed bearing in mind that all of them entail both strengths and weaknesses in regards to the way developed, implemented and data gathered. Nonetheless, these were adopted falling in the offset and completeness ways of combining qualitative and quantitative methods as identified by Bryman (2008a) and (2008b). As a result, a combination of methods was used to offset the weaknesses of the methods employed and draw on the strengths of both. Questionnaire was used as a mean to easily define users' characteristics but it could only provide hints on users' search behaviour, which could be further exploited during interviews. In addition, observation enabled the identification of specific search behaviours and actions on specific occasions and the formulation of relevant questions but without the users' explanations during retrospective thinking aloud and interviews would be meaningless. Moreover, the retrospective thinking aloud enabled the collection of data entirely on users' explanations of their actions and description of their thoughts in retrospect rather than on facilitator's observation. On the other hand, it was as a process exhausting for the users because each session lasted approximately an hour and a half. Finally, interviews though they were kept short in length provided the chance to verify further users' answers in the questionnaire and specific actions noted during observation and clarify users' expressions during retrospective thinking aloud.

The notion of completeness was also important in identifying users' information seeking behaviours and reflecting on the development of CLIR systems. Taking the view of Bryman (2008a) and (2008b) on completeness, we employed four different methods coming both from quantitative and qualitative methods in order to form a more comprehensive account of users' information seeking behaviour in multilingual environments. In particular, we were able to define our sample and discover hints on our users' searching behaviour of which to verify on retrospective thinking aloud and/or ask during interviews. In addition, observe users' behaviour and clarify their actions accompanied by their thoughts, expressions and justifications during retrospective thinking aloud and interviews. The adoption and combination of these four methods enable the generation and gathering of diverse data on different aspects of the same research area, users' image seeking behaviour in multilingual environments. The data that was gathered would have been extremely difficult if not impossible to gather by using any other method(s) alone or in combination.

In addition, this notion of completeness was reinforced by the specific order in which the four methods have been applied. It provided the facilitator with the possibility of gradually accumulating data on users' search behaviour. In the beginning there would be only hints about users' general search characteristics, then notes on specific actions and finally users' own words and explanations regarding their actions and thoughts. Finally, the facilitator would only need to form questions clarifying, verifying and justifying users' actions during interviews. Therefore, the data gathered was providing a full and comprehensive account of users' image seeking behaviour in multilingual environments. Data extracted based on users' own comments rather than facilitators observations and assumptions.

9. Conclusion

The investigation and monitoring of users' interaction and search behaviour with multilingual environments is considered to be one of the most important factors in developing efficient and effective CLIR systems. As a consequence, the decision on which methods (qualitative and/or quantitative) to employ in order to gain a better insight to users' behaviours and thoughts is also essential. This study

used a mixed methods research understanding that mixed methods enable better (stronger) inferences and can answer questions that single methods cannot. In particular, four methods both quantitative and qualitative have been adopted such as questionnaire, observation, retrospective thinking aloud and interviews. These methods have been designed and implemented to draw insights on different aspects of users' characteristics and/or search behaviours. The defined order of their implementation was used to gradually gain a better and more comprehensive insight into users' actions and thoughts and to enable the formulation of relevant questions. In addition, the offset and completeness way of combining the chosen methods have been adopted in order to offset their weaknesses and draw on their strengths enabling the better and a more comprehensive understanding of users' image seeking behaviour in multilingual environments.

References

- Bryman, A. (2008a). *Social research methods*. Oxford University Press, Oxford; New York.
- Bryman, A. (2008b). Integrating qualitative and quantitative research: how is it done?. In *The mixed methods reader*. V.L. Plano Clark and J.W. Creswell (eds). University of Nebraska- Lincoln, California.
- Campbell, D.T., and Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, Vol. 56, 81-105.
- Clough, P., Gonzalo, J., Karlgren, J., Barker, E., Artiles, J. and Peinado, V. (2008). Large-scale Interactive Evaluation of Multilingual Information Access Systems: the iCLEF Flickr Challenge. Workshop on *Novel Methodologies for Evaluation in Information Retrieval*, 30 March 2008, Glasgow, Scotland, pp. 1–5.
- Denzin, N.K. (1978). *The research act in sociology: a theoretical introduction to sociological methods*. Butterworths, London.
- Ericsson, K.A., and Simon, H. (1985). *Protocol Analysis: Verbal Reports as Data*. MIT Press, Cambridge, Mass. ; London.
- Haak, M., de Jong, M., and Schellens, P. J. (2003). Retrospective vs. Concurrent Think-aloud Protocols: Testing the Usability of an Online Library Catalogue. *Behaviour & Information Technology*, Vol. 22, No. 5, 339 – 351.
- Peinado, V., Gonzalo, J., Artiles, J. and Lopez-Ostenero, F. (2008). UNED at iCLEF 2008: Analysis of a Large Log of Multilingual Image Searches in Flickr. In Peters et al (eds): *Evaluating Systems for Multilingual and Multimodal Information Access 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers, Springer LNCS (to be published)*.
- Tashakkari, A. and Teddi, C. (2003). *Handbook of mixed methods in social and behavioral research*. SAGE, London.
- Vassilakaki, E., Johnson, F., Hartley, R. and Randall, D. (2008). A Study of Users' Image Seeking Behaviour in Flickling. In Peters et al (eds): *Evaluating Systems for Multilingual and Multimodal Information Access 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers, Springer LNCS (to be published)*.