Strangers in a Strange Land: A Study of Second Language Speakers Searching for e-Services

ABSTRACT

While the recent trend of digitisation of government and related services offers many advantages, it could introduce problems for those who are less information literate or who have particular issues searching for and understanding the necessary content. In this study ten participants, who speak English as a foreign language, were given four search tasks designed to reflect actual information seeking situations. They completed pre- and post-search questionnaires to identify the relevancy of the task, their English language ability and search experience.

Our results suggest that, despite a perception that they performed to the best of their abilities, were bookmarking relevant documents and that the given tasks were easy, the students were actually often choosing documents that are only partially or tangentially relevant. The repercussions of this discrepancy are clear and suggest that much more assistance is needed before such services can be made 'digital by default'.

Keywords

information behaviour, search, user studies, e-services, information literacy

1. INTRODUCTION

The digitisation of services and dissemination of Internet-based technologies have resulted in initiatives, such as the UK government's 'digital by default', whereby public sector services are transitioned onto electronic systems (e-services). For the average user, for whom Internet technologies have become an everyday tool, this will come as no great burden; however, information and digital divides place importance on the IT literacy and abilities of the user [4].

For users of such systems whose first language is not English, searching for information in a foreign language could raise numerous problems surrounding: knowledge of content in a unfamiliar culture and setting; language use, where variations in meaning can significantly change search results and awareness of the reliability of sources. Users may have previously relied on face to face encounters with staff or the knowledge and experience of friends, family or community members when their own was lacking. However, situations where these social groups may not be attainable or knowledge and experience is just as lacking can have dire repercussions for some members of society who already face significant barriers [10].

Conference'17 Date, place

DOI: http://dx.doi.org/10.1145/2766462.2767731.

2. RELATED WORK

A number of studies have looked at multilingual online searching with research focusing on web content (or lack thereof) in the user's native language [2] and language proficiency as major contributors to the user's decision to search in a foreign language, notably when English is the foreign language [2, 3, 11]. Research within the field goes some way to disclosing the search behaviours of multilingual searchers but focuses predominately on "why" rather than "how" they search and how well they perform [12].

In a study of refugee use of government services Lloyd et al. [10] found that information poverty was a product of the social exclusion of the participants as a result of barriers e-services can erect. With this in mind, this paper is motivated by the findings of Rozsa et al. [12] and Lloyd et al. [10] in observing the online search behaviours of non-native users of UK governmental e-services. It acknowledges the participants use of such services as they are now digitised (the 'why') and looks to compare the search behaviours of users whose first language is not English (the 'how').

3. METHODOLOGY

Identification of the types of services such users would use was made by the author in an (as yet) unpublished study in which 7 international PhD students at a UK university, 6 of whom also took part in the study described in this work, were recruited and were tasked with identifying: what a government service entailed; which would be deemed most useful (to the group); and the information needs, information sources and skills that would be required to successfully utilise the e-service. The findings from this study helped shape the task scenarios detailed within this paper.

The 10 study participants were all international PhD students from a large UK university who spoke English as a second language (fluent 80%, competent 20%). All participants were from different countries across Europe (20%), Asia (70%) and Africa (10%) with a total of 11 languages spoken natively, and 15 languages in total up to a competent level. 40% of the participants were female with an average

age of 31 ($SD = 3.56$) and 60% were male with an average age of 31.5 ($SD = 3.33$). Each was remunerated for their participation with a £10 Amazon voucher.

There were a total of three sessions with two sessions of three students and one of four as dictated by participant and technical equipment availability. Each session followed the same process of each participant filling in a demographic questionnaire which collected information on their area of study; age; gender; nationality; language(s) spoken and proficiency; IT use; search engine use in English and their native tongue; search engine competency and preference and their own UK governmental service awareness. The participants were then asked to perform four search tasks. Using the Chrome browser, each participant was asked to use the Google Search Engine to start each task but were not limited to the search results page.

Each task was allotted ten minutes for completion with up to five minutes for the participants to read the task and complete the pre- and post-questionnaires. This allowed for no more than one hour in total. Tasks were distributed to participants using a Latin square design to account for task fatigue and potential learning effects [9].

3.1 Tasks

The search tasks were defined based on the results of the previously mentioned short study and are designed to reflect actual information seeking situations in an attempt to be relevant and a more interesting search experience for the participants [6].

1. *Your friend from Peru and their family (2 members) are coming to visit you for 6 months while you are in the UK. Develop a list of instructions to help them apply for the necessary visas.*

2. *A family member is coming to the UK to live and wants information on housing. They have heard there are a number of options and have asked you for advice. Identify the options available to them and recommend which they should choose. Give reasons to support your recommendation.*

3. *Your friend just got back from a trip abroad and suddenly developed a high fever. A dry cough, chills, and breathing difficulties soon followed. What could they have? They have no insurance and have asked your advice on what to do. Provide them with recommended actions.*

4. *Your elderly neighbours have heard about the UK governmen'ts 'digital by default' initiative and are concerned about whether this will affect them and their friends at the local community centre. They have asked you to find out more about it. Use your best judgement to highlight what would impact them with reasons for your choices.*

All 4 tasks were assessed by the participants as being relevant or partially relevant to them with task 1 receiving the highest average relevance score and task 4 the lowest.

Prior to beginning each task, participants were asked to fill in a pre-task questionnaire [6] (see Table 1) to gauge their domain knowledge, interest in the topic and the perceived difficulty of the task using a five point Likert scale where 1 is "Not at all" and 5 is "Very".

Q1	I have searched about this topic before
Q2	I know about this topic
Q3	I am interested in this topic
Q4	It will be difficult to find information about this topic

Table 1: Pre-task questions.

As mentioned prior, the participants' activities were logged with video and audio data recorded and they were asked to bookmark any website (document) deemed relevant to the task. At the end of each task the participant was also required to complete a post-task questionnaire [1, 5, 6, 8, 9] with responses on a five point Likert scale where 1 is "Not at all confident" and 5 is "Very confident". Examples can be seen in Table 2.

Q3	The task was relevant to me
Q6	I performed the task to the best of my ability
Q7	I found the task difficult
Q8	I'm confident the content I found satisfied the task
Q10	I'm confident I identified relevant websites
Q11	I'm confident in my ability to read the website content
Q12	I am confident in my ability to understand the content of the websites I visited

Table 2: Selected post-task questions.

Afterwards participants were invited to take part in a semi-structured group discussion to examine their searching experiences and to highlight any issues or concerns and offer suggestions or solutions to any of the points that arose. Findings from these discussions are not included in this paper due to space limitations.

To determine the relevance of the bookmarks logged by the participants, all bookmarks were assessed by two IR researchers using a voting strategy and given scores between 1 and 4, where 1 is not relevant, 2 is tangentially relevant, 3 is partially relevant and 4 is totally relevant. Any bookmarks not assigned the same score by the two assessors in the first round were discussed and a single score was agreed, although this only occurred for a very small number of cases.

4. RESULTS

In total participants created 267 bookmarks, with an approximately equal split between governmental and non- governmental resources. Table 3 shows the total number of bookmarked URLs that participants deemed relevant to the tasks and whether they were from governmental sources and the mean relevance score. Of all the URLs bookmarked only 60.7% were either partially or totally relevant, with 30.7% tangentially relevant and 8.5% non-relevant and there were no significant differences between the median number of bookmarks per task with each task receiving 8 or 9 per participant on average. Surprisingly, there was little difference in terms of relevance between governmental and non-governmental resources. This was mostly due to some participants bookmarking internal policy documents or documents discussing best practices for civil servant software engineers which were deemed to be only tangentially relevant and unlikely to be of help in the given contexts.

English proficiency was self-assessed with 80% of the participants declaring themselves fluent and 20% competent

Bookmark	No.	Relevance
Governmental:	141	2.91
non-Governmental:	129	2.92

Table 3: Bookmark types

with all participants using IT daily and formulating queries (on search engines) in English on a daily basis (90%) or a few times a week (10%). When judging their own abilities in formulating queries in English, identifying relevant search results and information on websites, 8 out of 10 were either confident or very confident in their abilities with the remaining 2 stating they were "slightly confident". Half of the participants had used UK government e-services previously, 30% hadn't and 20% were unsure what was meant by the term.

As shown in Table 4, there was considerable variation in performance by different users with the bookmarks of 5 participants being only relevant in 50% or less of cases. There was also variation in the numbers of pages bookmarked; one participant only bookmarked 3 per task on average with the majority bookmarking 5 or more. Participant J acknowledged their limited bookmarks for the 3rd task as in a real scenario they would not risk the health of another by self diagnosing, and would instead only refer that person to a health professional in the first instance.

User	Precision	Bookmarks/task
A	0.81	6
B	0.74	11
C	0.50	7
D	0.43	10
E	0.44	9
F	0.43	9
G	1.00	6
H	1.00	3
I	0.80	5
J	0.49	9

Table 4: User performance.

When comparing the performance of the participants against the perceived task difficulty (pre-task question 4) they were unable to successfully predict how well they would perform (Figure 1 and Table 5), in fact it appears that they performed best in cases where they expected to perform poorly! In spite of this, there is some correlation, albeit not significant (0.26, p-value = 0.11), between their post-task assessment of difficulty and their pre-task prediction of the same. This suggests that they were unlikely to dramatically change their perception of a task's difficulty after having actually performed that task.

Confidence in their own abilities was in evidence throughout the study and, even in the case of the few who doubted their abilities in the pre-study questionnaire, the level of confidence was predominately in the confident to very confident response levels. As shown in Table 6 and Figure 2, responses to the post-task questions on self-perceived performance (Q6), confidence in having completed the task well (Q8) and of having identified relevant websites (Q10) were overwhelmingly positive. This was also the case with regards to understanding what they had read during the task (Q11)

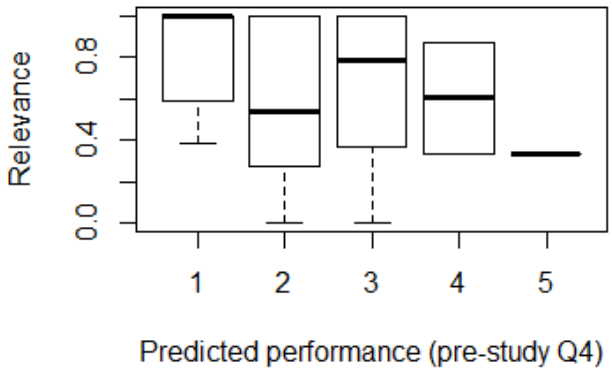

Figure 1: Performance by expected task difficulty.

Pre-study Q4	Relevance	Count
1	1.00	7
2	0.53	18
3	0.79	12
4	0.60	2
5	0.33	1

Table 5: User performance vs expected task difficulty.

and on the websites bookmarked (Q12). This is further reflected in their categorisation of task difficulty (Q7), where the median response was 2 (not difficult). This confidence is in stark contrast to their overall performance, which was generally quite poor.

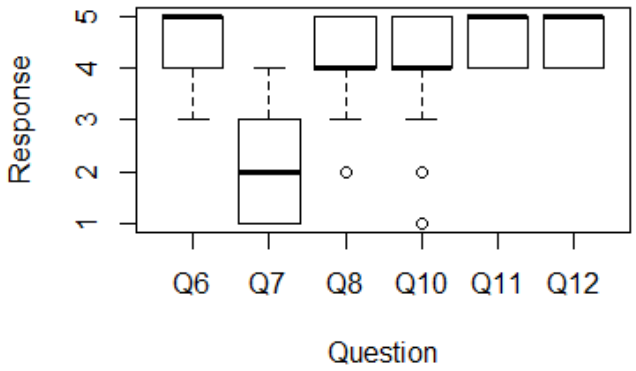

Figure 2: Confidence in abilities and task difficulty.

Further to their views of their abilities, the post-task review of their perceived performance in comparison to their actual performance shows participants were not able to correctly determine how well they had performed. The majority of responses were in the confident to very confident range while precision was actually lower than for the less confident ratings ($p < 0.76$) as shown in Table 7.

Response	Q6	Q7	Q8	Q10	Q11	Q12
1	0	14	0	1	0	0
2	0	10	3	1	0	0
3	4	9	3	1	0	0
4	7	7	18	21	17	16
5	29	0	16	16	23	24

Table 6: Confidence in abilities and task difficulty.

Response	Performance		Count	
	Q8	***Q10***	***Q8***	***Q10***
1	0	1.0	0	1
2	1.0	0.22	3	1
3	1.0	1.0	3	1
4	0.61	0.63	18	21
5	0.69	0.69	16	16

Table 7: Confidence in content that satisfied the task (Q8) and confidence in relevant websites from search results (Q10) versus performance.

5. CONCLUSIONS AND FUTURE WORK

In this work we have investigated the problem of people needing to search for, and assess, documents in a language foreign to them. Two factors contribute to this becoming an ever more common situation: the increasing numbers of people working and studying abroad as well as those who arrive in countries as refugees; and the push by many governments and organisations to move to fully-digital systems, the so-called 'digital by default' or 'e-government' movements.

The results of our 10-participant user study suggest that this situation may result in people overestimating their abilities and assessing non-relevant sources as being relevant and helpful. Our participants found the tasks to be relevant, assessed their English-language searching and reading abilities to be good and, having completed each task, were confident in their search and bookmarking performance. In spite of this, almost half of the documents they selected were assessed by native speakers to be either non-relevant or only tangentially relevant. The potential repercussions of this, had these situation been real, could be quite costly. Is this due to cultural differences, domain knowledge, or do they rely too much on government/non-government materials. Understanding why these issues arise may inform development of systems and techniques to mitigate these problems, perhaps by "training" users to search better via instruction or "learning by example" [7].

This paper only touches on the findings of this study and we intend to build on this work through further investigation into the online search behaviours of users who speak English as a second language. In future work we intend to use the various sources of data collected (including screen capture) to investigate the search behaviours and strategies participants employed. Using appropriate software we can mark all of the actions performed by participants during the sessions and calculate metrics such as the time taken to generate queries, the number of query reformulations, query terms used, the amount of time spent on SERPs and how and where the bookmarked documents were actually found. We intend to run the same study with native speakers to determine whether their performance is indeed better, as one would expect and hope, and compare them with the non-native speakers.

6. REFERENCES

[1] D. Bell and I. Ruthven. Searcher's assessments of task complexity for web searching. In *Advances in Information Retrieval*, pages 57–71, Berlin Heidelberg, April 2004. Springer.

[2] B. Berendt and A. Kralisch. A user-centric approach to identifying best deployment strategies for language tools: the impact of content and access language on web user behaviour and attitudes. *Information Retrieval*, 12(3):380–399, 2009.

[3] T. Bogers, M. Gäde, M. Hall, and M. Skov. Analyzing the influence of language proficiency on interactive book search behavior. *iSchools*, 2016.

[4] P. Chu, E. Jozsa, A. Komlodi, and K. Hercegfi. An exploratory study on search behavior in different languages. In *Proceedings of the 4th Information Interaction in Context Symposium*, pages 318–321. ACM, 2012.

[5] K. Collins-Thompson, S. Rieh, C. Haynes, and R. Syed. Assessing learning outcomes in web search: A comparison of tasks and query strategies. In *Proc. of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, pages 163–172. ACM, March 2016.

[6] A. Edwards and D. Kelly. How does interest in a work task impact search behavior and engagement? In *Proc. of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, pages 249–252. ACM, March 2016.

[7] M. Harvey, C. Hauff, and D. Elsweiler. Learning by example: training users with high-quality query suggestions. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 133–142. ACM, 2015.

[8] B. Johnston and S. Webber. Information literacy in higher education: a review and case study. *ACM Trans. Program. Lang. Syst.*, 15(5):795–825, November 2003.

[9] D. Kelly, J. Arguello, A. Edwards, and W. Wu. Development and evaluation of search tasks for iir experiments using a cognitive complexity framework. In *Proc. of the 2015 International Conference on The Theory of Information Retrieval*, pages 101–110. ACM, September 2015.

[10] A. Lloyd, M. A. Kennan, K. M. Thompson, and A. Qayyum. Connecting with new information landscapes: information literacy practices of refugees. *Journal of Documentation.*, 69(1):121–144, 2013.

[11] J. Marlow, P. Clough, J. Recuero, and J. Artiles. Exploring the effects of language skills on multilingual web search. In *Advances in Information Retrieval*, pages 249–252, Berlin Heidelberg, March 2016. Springer.

[12] G. Rózsa, A. Komlodi, and P. Chu. Online searching in english as a foreign language. In *Proceedings of the 24th International Conference on World Wide Web Companion*, pages 875–880. International World Wide Web Conferences Steering Committee, 2015.