The effect of Dyslexia on Information Retrieval: a pilot study

A. MacFarlane[1], A. Al-Wabil[2], C. R. Marshall[3], A. Albrair[1], S. A. Jones[1] and P. Zaphiris[2]
Centre for Interactive Systems Research[1]
Centre for HCI Design[2]
Department of Language and Communication Science[3]
City University London, Northampton Square, London EC1V 0HB

Abstract

Purpose – The purpose of the paper is to resolve a gap in our knowledge of how people with dyslexia interact with Information Retrieval (IR) systems, specifically an understanding of their information searching behaviour. Very little research has been undertaken with this particular user group, and given the size of the group (an estimated 10% of the population) this lack of knowledge needs to be addressed.

Design/Methodology/Approach - We use elements of the dyslexia cognitive profile to design a logging system recording the difference between two sets of participants: dyslexic and control users. We use a standard Okapi interface together with two standard TREC topics in order to record the information searching behaviour of these users. We gather evidence from various sources, including quantitative information on search logs, together with qualitative information from interviews and questionnaires. We record variables on queries, documents, relevance assessments and sessions in the search logs. We use this evidence to examine the difference in searching between the two sets of users, in order to understand the effect of dyslexia on the information searching behaviour. A topic analysis is also conducted on the quantitative data to show any effect on the results from the information need.

Research limitations/implications – As this is a pilot study, only 10 participants were recruited for the study, 5 for each user group. Due to ethical issues, the number of topics per search was restricted to one topic only. The study shows that the methodology applied is useful for distinguishing between the two user groups, taking into account differences between topic. We outline further research on the back of this pilot study in four main areas. A different approach from the proposed methodology is needed to measure the effect on query variables, which takes account of topic variation. More details on users are needed such as reading abilities, speed of language processing and working memory to distinguish the user groups. Effect of topic on search interaction must be measured in order to record the potential impact on the dyslexic user group. Work is needed on relevance assessment and effect on precision and recall for users who may not read many documents.

Findings – Using the log data, we establish the differences in information searching behaviour of control and dyslexic users i.e. in the way the two groups interact with Okapi, and that qualitative information collected (such as experience etc) may not be able to account for these differences. Evidence from query variables was unable to distinguish between groups, but differences on topic for the same variables were recorded. Users who view more documents tended to judge more documents as being relevant, either in terms of the user group or topic. Session data indicated that there may be an important difference between the number of iterations used in a search between the user groups, as there may be little effect from the topic on this variable.

Originality/Value – This is the first study of the effect of dyslexia on information search behaviour, and provides some evidence to take the field forward.

Keywords: information retrieval; dyslexia
Article Type: Research paper

Introduction

While the existence of dyslexia has been recognised for more than 100 years, it is only recently that efforts have been made in the research community to understand the effect of this cognitive disability, and to support people with the condition in their daily lives (Snowling, 2000). Dyslexia for the purpose of this study is defined as "A learning disorder marked by impairment of the ability to recognise and comprehend written words", a standard definition found on many web sites. Most work has been in the area of child pedagogy (Snowling, 2000), although there is increasing interest in the problems of adults (Morgan and Klein, 2000). There is a paucity of research in the information retrieval community for either type of user. There has been some work in the area of assisting users with query errors (Sitborn et al, 2007) and the effect of font type and line length on browsing (Ling and van Schaik, 2006), but there appears to be little or no work in the area of information seeking and searching. In this study we intend to address the lack of research in this area by conducting a real world user evaluation comparing dyslexic and non-dyslexic users, in order to understand the effect of dyslexia on the information searching process. The paper is structured as follows. We motivate the study, and then define dyslexia and its potential effect on information retrieval (IR) given the cognitive profile of users who have been diagnosed with the condition. Related work is then described. We then set out our research hypotheses given important elements of the cognitive profile. The participants and experimental methods are described followed by an analysis of the data collected. The implications of the results from the study are then discussed. A conclusion is given at the end, together with further work required in the area.

Motivation for the study

According to the British Dyslexia Association the estimated proportion of the UK population who have dyslexia is around 10% (BDA, 2007), which is divided into those people who are severely dyslexic (4%) and those people who are mildly dyslexic (6%). For the UK alone, with a population of 60 million, this gives us an estimate of around 6 million people. If we take this estimate and apply it to the European Union (population around 350 million) and the United States (250) million, this yields an estimate of around 60 million people who have some kind of problem due to a form of dyslexia in a part of the developed world. This is too large a user group (the size of the UK) to ignore. There is a clear gap in the literature concerning the behaviour of people with dyslexia when they seek and search for information, and our primary aim in this research is to inform the community of the impact of this condition. As this is a pilot study, our long term goal is to provide information on how to improve the experience of dyslexic users' experience of searching for information.

The dyslexia cognitive profile and Information Retrieval (IR)

Dyslexia is classed as a specific learning difficulty, an umbrella term which includes conditions such as dyspraxia (motor difficulties) and dyscalculia (mathematics difficulties). With dyslexia the learning difficulty is in reading words (the etymology of the term is Greek, *dys* meaning "problematic" and *lexis* meaning "words"). More specifically, dyslexia is a problem with linking phonemes (the sounds of a language) and graphemes (e.g. the letters or characters used to represent those sounds in written form). Dyslexia is a difficult condition to diagnose, principally because it is identified through exclusionary criteria. It is diagnosed in individuals who have severe difficulties in learning to read and spell despite having no obvious verbal or non-verbal impairment, sensory deficit (e.g. a visual deficit or hearing loss), pervasive developmental disorder or frank neurological impairment (Snowling, 2000). Moreover, the disorder is not static – people and circumstances change, and this may affect the performance a person on a given task (for example, searching for information) over time.

Behaviourally, dyslexia affects reading and spelling. In terms of IR, this means that users may not be able to recognise familiar words, may find it difficult to attempt to decode unfamiliar words, may make mistakes when spelling words, and may read and type more slowly than is the norm. The impact of this is easy to imagine. Such users are predicted to have greater problems than non-dyslexic users in formulating a useful query. They may also have problems identifying useful terms from documents, and hence identifying relevant concepts to be used to refine their searches. They may have difficulty in applying query modification techniques, particularly those of the manual variety.

Cognitively, deficits in phonological processing, and particularly in reduced working memory capacity for written material, are proposed to underlie dyslexia (Morgan and Klein, 2000). Working memory problems could have a significant impact on IR – if a user is unable to retain the necessary information in one part (or iteration) of a search session, the implication is that he or she will not be as effective a searcher as a non-dyslexic user (at least within a single iteration). Furthermore, dyslexic people have problems with sequencing, and this is predicted to have an impact on searching behaviour because the process of searching is a sequential process.

The learning style of dyslexics can be summed up as 'obsessional and labour intensive'. Do these users keep up the IR process above and beyond that of a non-dyslexic user? Will they, for example keep refining their query and use more iterations, and will they look at more screens of results (rather than just one)?

Related work

There is much interest in dyslexia within the field of Human Computer Interaction (HCI), focusing on the problems dyslexic users have with issues such as the layout and structure of interfaces (Gregor et al, 2003; Kurniawan, S. et al, 2006) and web interaction (Alty, 2002; Sackville et al, 2002). The Disability Rights Commission published a major investigation of access to the web for various types of disabled persons including dyslexic users (DRC 2004). The study included a set of tasks which involved interacting and navigating web sites using a set of pre-defined tasks. A number of recommendations were made on the back of this study, including the improvement of search design (although little detail is given as to why this should be the case). More recently Al-Wabil (2009) conducted a study on the effect of dyslexia on Web navigation. Several areas were investigated in this study. Preliminary investigations established the barriers of this user group to web access, and used eye-tracking methodology to investigate differences in visual attention differences between control and dyslexic users. Two further experiments were conducted in order to extend those preliminary investigations. Investigating dyslexic users' function problems when interacting with web structures revealed that dyslexic users took longer to complete tasks and experienced more disorientation on web structures than control users. Evidence demonstrated that navigation problems were associated with elements of the cognitive profile mentioned above e.g. reduced working memory capacity and phonological processing difficulties, as well as visual stress. A second experiment found a difference in scanpaths (paths of eyes when scanning pages) between controls and dyslexic users when viewing text heavy pages. Whilst that study focused on search by undirected browsing, the implications for this study – focused on directed searching – are profound.

Research question

The research question to be tackled in this paper is to establish the differences in information searching behaviour between control and dyslexic users, as measured by users' ability to interact with the system. We have some evidence from general web search searching behaviour studies (Jansen et al, 1998; Silverstein et al 1999; Spink et al, 2002) indicating how this research question can be examined, which helps to narrow our focus:

- Query size: 2-3 for web users
- Query modification methods: manual, automatic
- Page views: web users only look at 1st page for the most part (and only the top set of results at that).
- Type of logic: Boolean, proximity, natural language (implicit operators)

This evidence is used to examine searching behaviour with a focus on search sessions, including terms for queries and their size, modification of those queries, reading and making relevance assessments on documents, and examining result lists of documents. This gives us three broad areas of variables to examine: query, document and session based. With regard to reading and spelling factors we test the ability of the users to enter mistake-free queries on the initial search, and their subsequent reading abilities throughout the search interaction by viewing documents. In terms of memory we examine the interaction with sessions (search iterations within a search), results lists and relevance pools (a set of documents assessed to be relevant to a given topic). This information leads directly to the experimental design described in the next section, and leads to a secondary research question – is this methodology useful for the proposed research agenda?

Participants and methods

In this section we outline our experimental methodology and data collection methods for search. These are based on a design first published in MacFarlane et al (2005), presented and discussed at the COLIS 2005 user studies workshop. That paper described the experimental design and did not contain any results. The design is focused on collecting the query, document and session variables identified in the research question above. We also use pre-search questionnaires to gather some details about the users and to ensure they are in distinct groups: either control or dyslexic user. Post search interviews are then used to capture evidence about the search experience. We describe the participants of the study in this section, experimental procedure and apparatus used, the pre-search questionnaire, information about the search interface, the test collection used, search data recorded in logs and finally the post search interview and debriefing based on the requisite experimental design.

Participants

Data	Control Users	Dyslexic Users
Male	2	2
Female	3	3
Age Range	23-33	23-52
Education attainment	1 PhD, 1 MSc, 1 MA, 1 BSc and 1 undergrad student	1 PhD, 1 MSc, 1 BA and 2 undergrad students
Occupation	1 Undergrad student, 2 Graduate students, 1 Programmer and 1 Research Fellow	2 Undergrad students, 2 Graduate students and 1 Lecturer
Computer usage	10-22 years	6-25 years

Table 1 – Details of participants

We recruited a total of 10 participants for the study, 5 control users and 5 dyslexic users. Their details, including their age range, are given in table 1. Participants are either highly educated or are undergraduates. Most participants are students. Each participant was given an information sheet detailing the purpose of the project, the procedure and explaining the long term benefits of the study. Participants were remunerated with a gift voucher of £25 for their time and expenses. Throughout this paper, dyslexic users are labelled D[1-5] and control users labelled are C[1-5].

Participants in this study had been involved in previous studies (Al-Wabil et al, 2008; Al-Wabil, 2009) discussed in the related work section. Those allocated to the dyslexia group all had a diagnosis of dyslexia from an Educational Psychologist. None of the control participants had a diagnosis of dyslexia.

To confirm group membership we asked all participants to fill in the British Dyslexia Association's checklist (Vinegrad, 1994). This is a standard self-assessment questionnaire that people can use prior to diagnosis, in order to find out whether it is worth their while having a fuller assessment for dyslexia. It is simple and quick to use, requiring only a few minutes to complete. There are 20 questions, and each requires just a "yes" or "no" response. 9 or more "yes" responses could indicate that the respondent has dyslexia. The reader is referred to (BDA, 2009) for details of the questions in the checklist.

ID	Number of "yes" responses/20
D1	10
D2	16
D3	11
D4	18
D5	15
C1	0
C2	1
C3	2
C4	2
C5	1

Table 2 – Individual scores on BDA Checklist

Checklist data are shown in Table 2. The subjects have clearly put themselves in two distinct groups using the checklist. The average BDA scores for the control user group is 1.2 compared with a dyslexia user group figure of 14.0.

Experimental procedure and apparatus

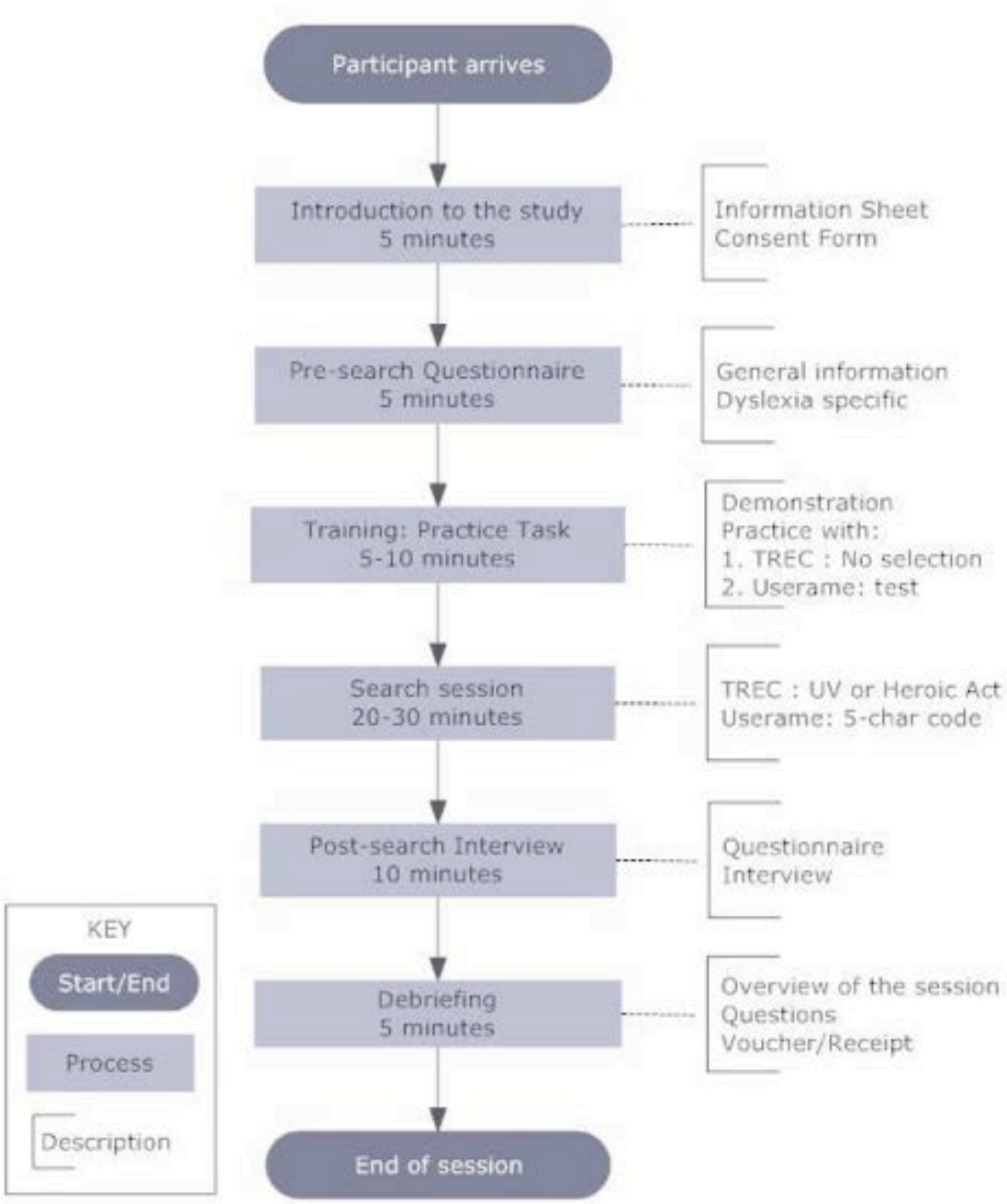

Figure 1 – Experimental procedure

The experimental procedure is shown in Figure 1. On entering the Interaction Lab, the participant was asked to take a seat in front of the participant PC and monitor, at a distance of approximately 60cm from the monitor. The study was explained to the participant, including an overview of the interface, the purpose of the study and the nature of the search task to be conducted in the session. The information sheet was administered, followed by the consent form. The session began with the administration of the pre-search questionnaire. The interface was then demonstrated by the experimenter. The procedure for logging in to the IR system, searching using Query terms, modifying Query Terms and judging the relevance of documents was explained until the

participant indicated understanding of this. Users completed this procedure only once – no retests were undertaken.

Pre-search questionnaire

General information	
1) Demographic information: What is your age? What is your sex? 2) Knowledge of search engines/online services: What search engines do you use? Do you use any online services e.g. Dialog/Factiva?	3) Knowledge of resources: online, hardcopy, web: What papers/magazines do you read? How would you define information quality? 4) Knowledge of computers and the internet: How often do you use the internet? Do you have a broadband connection? 5) How often do you do searching?

Dyslexic specific	
6) Have you been diagnosed as dyslexic? Yes (finish questions) No (Go to Q7) 7) Do you think you are dyslexic? Yes (go to Q8) No (finish questions) 8) Do you have any difficulties when reading silently? (yes/no) Yes No 9) If yes, does it involve any of the following phenomena? Words move around on the page. Words disappear from the page. Spaces between words form "rivers" down the page 10) Do you have any difficulties when reading aloud? Yes No 11) Do you have difficulty spelling words? Yes No	12) Do you have difficulty thinking in a "linear" manner? Yes No 13) Would you say you were more than averagely clumsy? Yes No 14) Do you confuse your left and right hands/side sometimes? Yes No 15) Do you have difficulty navigating (either in the real world or in virtual worlds such as the World Wide Web)? Yes No

Table 3 – Pre-search questionnaire

The pre-search questionnaire (see table 3) was used to gather general information about the user, and the user's own impression of their dyslexia (in the case of controls the lack of the condition). In the general information section we collected standard data about the age and sex of the participants (presented above), as well as information on their knowledge of search, sources and quality in order to gauge their prior knowledge of information retrieval. In the dyslexia segment we pick up specific issues which may affect the users undertaking the study, particularly as we did not have access to the dyslexic users' assessment reports.

Search time data collection: Okapi interface, test collection and logging

Figure 2 – Okapi interface

We used the Okapi system for our experiments, utilising a PHP web interface developed for conducting user experiments (Vakkari et al, 2004) – see Figure 2. The interface consists of three frames: one to enter the initial query and provide various query and document options, one to display the current search terms and one to display documents lists, and documents. The user initiates the search by typing in search terms and pressing *return* – this puts an initial list in the search term frame. Clicking on the *search* button retrieves a hit-list of documents, which the user can view on request. Users can delete terms from the Query terms frame at any time. When three documents have been judged to be relevant, the user can then press the *expand* button to initiate relevance feedback – once expand has been requested no more search terms can be entered into the search box. This strategy was used to see how far the user would go with automatic and manual relevance feedback techniques (the latter being the deletion of terms). Users can view a list of documents judged relevant on request (by clicking on *view your choices* button). When finished the user presses the *exit* button to complete the session.

A logging mechanism was used to record the variables we examine in relational tables - see figure 3. Sessions are recorded from the moment the user logs in until the end of the session (user clicks on the *exit* button). The variables for the experiments are split into three groups identified in the research question section above: terms, documents and sessions. These variables are used to tackle our stated research question - to examine the difference between control and dyslexic users.

i) Terms

- Query terms.
- Query size (number of user entered terms).
- Rate of spelling mistakes (%).
- Number of terms deleted in Relevance Feedback.
- Query exhaustivity: % of query terms which are expressed in facets.

ii) Documents

- Total documents examined.
- Number of Documents judged relevant.
- Number of Documents judged non-relevant.
- Number of Documents examined per iteration.
- Total changes of judgements from relevant to non-relevant.
- Level of agreement with TREC relevance judgements (%)

iii) Sessions

- Session length (secs).
- Number of searches (or search iterations).
- Number of expansions.
- Number of Hit-lists examined per iteration.
- Number of Pool views (Documents judged relevant) per iteration.

Figure 3 – Log information recorded in database (quantitative variables for study)

We use disk 4 and 5 of the TIPSTER collection from TREC 7 and 8. This consists of documents from the Financial Times, Federal Register, Foreign Broadcast Information Service and Los Angeles Times. The collection of documents is made up largely of news articles together with US Government documents. We use only two of the general topics from the test collection, namely 427 and 442, the facets generated in Vakkari et al (2004) for each topic and the number of subjects in each group per topic – see table 4. Users were asked to search for one topic only, fewer than many other interactive studies, e.g. Vakkari et al (2004) asked participants to search for four topics. Whilst this is problematic (topic has an effect), there are valid ethical reasons for restricting the workload on dyslexic users. Dyslexic people have a tendency to become tired after a period of concentration, and to restrict the possibility of discomfort for these participants, it was decided that all users only be asked to search one topic. This strategy was recommended by senior members of the IR community when discussing the experimental design at the COLIS workshop (MacFarlane et al, 2005).

No	Topic description	Facets	Controls	Dyslexics
427	Find documents that discuss the damage ultraviolet (UV) light from the sun can do to eyes.	• Condition: Eye damage, diseases, cataracts, ocular melanoma • Causes: Sun, UV, ultraviolet light	3	2
442	Find accounts of selfless heroic acts by individuals or small groups for the benefit of others or a cause.	• Activity: Heroic acts (particular) • Person: Individuals, small groups	2	3

Table 4: Topic descriptions and facet analysis

Topics were assigned to users on a round robin basis. With the control users the bias is towards topic 427, with dyslexic users the bias is towards topic 442. This is a factor of having an odd number of participants per group. However each topic had an equal number of searchers – 5 per topic. Results are examined on two levels: control vs. dyslexic users and per topic – the latter to provide some insight into the effect of the topic on the results. Eye tracking data were also collected during the search, but this will be reported separately for space reasons.

Post search interview and debriefing

The purpose of the interview and debriefing is to collect qualitative data on the users' experience with the Okapi interface and the search session. The open-ended questions used to collect this information are listed in table 5:

Focus	Questions
Usability of the interface	1. What did you like about the interface? 2. What did you dislike? 3. What feature would you like which was missing from the interface?
Satisfaction of search on topic	1. How difficult did you find the topics 2. Did you manage to find documents which satisfied your need?

Table 5 - Post search interview and debriefing questions

These data will be used to analyse the usefulness of the interface for our experiments. The questions are split into two main groups: usability of the interface and satisfaction of search on the topic given to the user. Questions on usability of the interface were used to elicit information on the effect of the interface on searching behaviour, while the questions on the search topic were used to elicit information on the user experience searching on the given topic in terms of relevance assessment.

Experimental results

Results from the pre-questionnaire are discussed in the first instance. Results of the actual experiments found in the search logs are discussed next, and then the results from post search questionnaires are tackled.

Pre-Search Questionnaire results

Table 6 show a summary of the results from the pre-search questionnaire. With regard to search engines use it should come as no surprise that Google (the search engine and other services such as Google Scholar, image and map search) is used by all users. The dyslexic user group is a little more varied in their use of search engines, using Yahoo, Copernic and Ask. Both groups use library catalogues, online services and various access points to academic articles. All users are clearly geared up for academic research. All participants read a wide range of newspapers and magazines with some interest in academic journals, free newspapers and leisure magazines. Both sets of users obtain information from a wide variety of sources. With regard to information quality, relevance to the information need and sources are mentioned most frequently, with some interest in content of the article e.g. clarity, argument building and conciseness and authorship. Both user groups have a sophisticated understanding of information quality and how to apply the concept during searching. There does not appear to be any substantial difference between the

groups on the information collected here - any differences between searching behaviour cannot be accounted for by the knowledge of search services and/or information quality.

Questions	Dyslexic User Group	Control User Group
2. Knowledge of search engines/online services		
What search engines do you use?	D1: Google (images, maps, Web search). D2: Google (Web search, Scholar), Yahoo, and Copernic. D3: Google. D4: Yahoo and Google (Yahoo more). D5: Google, Yahoo and Ask.	C1: Google. C2: Google and Microsoft Live search. C3: Google (Web search, Scholar and maps) C4: Google (Web search and Scholar). C5: Google.
Do you use any online services e.g. Dialog and Factiva?	D1: University Library catalogue. D2: ERIC Education Resources Information Centre. D3: Athens to access academic articles. D4: Factiva, British Library catalogue. D5: Athens for legal case research.	C1: LexisNexis, Lawtel C2: Microsoft Developer Network (MSDN). C3: ISI Web of Knowledge, journal websites, City Library and the British Library sites. C4: JSTOR, MedHist, and journal catalogs for arts and humanities. C5: ACM DL, IEEE Xplore, CU lib service (uses Google search to point to these articles).
3. Knowledge of resources: online, hardcopy, web		
What papers/magazines do you read?	D1: Web: BBC, Guardian, Telegraph. D2: Web: Guardian/New York times, both online. D3: Web: Economist, FT. Magazines: IJGIS, Environmental planning. D4: Papers: Metro, London Paper. Web: Times, CNN, and Guardian. D5: Papers: Telegraph/Sun.	C1: Papers: Times. Magazines: Now. C2: Papers: Guardian, Observer. Magazines: PC Pro, When Saturday Comes. C3: Web: BBC/Spiegel sites. Magazines: sports (occasionally). C4: Papers: Guardian (weekends only). C5: Papers: Metro/London Lite. Web: Google News, Gamespot.
How would you define information quality?	D1: Source, relevance. Web, addresses of results before clicking. D2: Look at publisher, bibliography, footnotes, academic or well known press, and year. Relevance. D3: Relevance, source, building argument in article. D4: General: Clarity, presentation and writing. Academic: references/citations. News –reliability of source. D5: Information source, relevance	C1: Depends on source. LexisNexis very good. General articles, date and relevance to need. C2: Written to present data clearly, not obscured with unfamiliar words. Short e.g. Slashdot. C3: Leisure: entertaining, surprise. Academic: author institution, Journal, relevance. C4: Clarity. Quality of argument, salient points, easily accessible C5: Leisure authors opinions when adds value, academic citations references to proper research. Politics neutral information.

Table 6 – Qualitative results from pre-search questionnaire (General Information)

Table 7 show the results on quantitative general information from the questionnaire. All users have a broadband connection. All users have substantial experience in using the Internet, with over 8 years of usage on average for dyslexic users and nearly 10 for the control group. There is not much difference between the groups in these data. Dyslexic users on average tend to use the Internet more hours per week, but the variation is greater due to one outlier (one dyslexic user declared they used the internet for 50 hours p/w). More control users tend to do searching more frequently than dyslexic users.

Question	D Users	C Users	D SD	C SD
4. Knowledge of computers and the Internet				
Experience of internet (in years)	8.4	9.8	3.65	1.10
How often do you use the Internet (hrs p/w)?	23.2	13.6	17.28	4.62
Do you have a broadband connection?	5/5	5/5	-	-
5. How often do you do searching?	2 M, 3 F	3 M, 2 F	-	-

Table 7 – Quantitative results from pre-search questionnaire (General Information) [M=many searches, F=few searches]

Table 8 shows the quantitative results for dyslexic-specific information. There is a clear difference between the users on being diagnosed with dyslexia and identification with the condition - all control users answered 'no' to most questions apart from a couple of instances. With regard to problems with reading (either aloud or silently) the results were the same. Navigation was a clear problem for all dyslexic users in all contexts, with only one control user declaring problems with navigating in the real world. Three dyslexic users declared more serious aspects of the condition such as words moving around the page and spaces forming rivers – only one user recorded problems with words disappearing (D2). Four dyslexic users have problems with spelling or confuse left and right, and three had problems thinking linearly. Most users did not feel that they had problems with clumsiness, but one dyslexic user had also been diagnosed with dyspraxia.

Question	**D Users**	**C Users**
6. Have you been diagnosed as dyslexic?	5/5	0/5
7. Do you think you are dyslexic?	5/5	0/5
8. Do you have any difficulties when reading silently?	3/5	0/5
9. If yes, does it involve any of the following phenomena:		
Words move around on the page	3/5	N/A
Words disappear from the page	1/5	N/A
Spaces between words form "rivers" down the page	3/5	N/A
10. Do you have difficulties when reading aloud?	5/5	0/5
11. Do you have difficulty spelling words?	4/5	0/5
12. Do you have difficulty thinking in a "linear" manner?	3/5	0/5
13. Would you say you were more than averagely clumsy?	2/5	1/5
14. Do you confuse your left and right hands/side sometimes?	4/5	0/5
15. Do you have difficulty navigating (either in the real world or in virtual worlds such as the World Wide Web)?	5/5	1/5*

Table 8 - Quantitative results from pre-search questionnaire (Dyslexia Information) – Note: * signifies problem only in real world.

In summary there is no strong evidence of difference between groups on the general information the users recorded, and they all have a clear understanding of searching and how to evaluate the quality of information in terms of relevance and an understanding of sources. Both sets of users have a lot of experience using computers and the Internet, and while control users show more evidence of searching, the difference is not large. However there is a clear difference between the groups when we turn to the dyslexic specific information declared in the questionnaire. All our dyslexic participants had a formal diagnosis of dyslexia, but this information, together with the information presented in the section on screening, confirms that they were still dyslexic at the time of our study.

Search log results

Measure	**D Avg.**	**D SD**	**C Avg.**	**C SD**	**% Diff over**

					controls
Terms					
Query size (user entered terms)	3.73	1.42	3.80	1.48	-1.75
Rate of spelling mistakes	0	0	0	0	0
Number of terms deleted in Relevance Feedback	11.8	16.8	8.66	5.28	36.2
Exhaustivity: % of query terms which are expressed in facets	90.00	22.4	70.00	27.4	28.6
Documents					
Total Documents examined	27.4	12.2	43.8	15.1	-37.4
Documents judged relevant	12.6 (46%)*	8.73	22.8 (52%)*	20.7	-44.7
Documents judged non-relevant	14.8 (54%)*	9.04	21.0 (48%)*	11.4	-29.5
Documents examined per iteration	8.32	4.51	6.42	1.84	29.6
Changes of judgements from relevant to non-relevant	0.40	0.55	0	0	N/A
Level of agreement with TREC relevance judgements (%)	48.5%	31.7%	68.0%	29.0%	-28.7%
Sessions					
Session length (secs)	1668	373	1752	226	-4.8
Number of searches	3.80	2.17	6.80	1.30	-44.1
Number of expansions	2.80	2.39	5.80	1.30	-51.7
Hit-lists examined per iteration	1.84	1.02	2.27	0.68	-18.9
Pool views (Documents judged relevant) per iteration	1.20	1.10	2.00	1.73	-40.0

Table 9 - Search log result averages and standard deviations (Dyslexic vs. Control users). (*percentage of total)

Measure	**442 Avg.**	**442 SD**	**427 Avg.**	**427 SD**	**% Diff over topic 427**
Terms					
Query size (user entered terms)	2.74	0.92	4.80	0.48	-42.9
Number of terms deleted in Relevance Feedback	14.0	14.8	6.41	2.73	118
Exhaustivity: % of query terms which are expressed in facets	70.00	27.4	100.00	0.0	-30
Documents					
Total Documents examined	35.8	21.8	28.2	10.7	26.2
Documents judged relevant	26.6 (62%)*	19.1	8.8 (31%)*	3.0	202.3
Documents judged non-relevant	16.4 (38%)*	9.40	19.4 (69%)*	11.9	-15.5
Documents examined per iteration	7.69	4.01	5.99	2.52	28.4
Changes of judgements from relevant to non-relevant	0.20	0.45	0.20	0.45	0.0
Level of agreement with TREC relevance judgements (%)	38.1%	24.7%	78.4%	21.4%	-51.3%
Sessions					
Session length (secs)	1839	295	1580	255	-16.4
Number of searches	5.40	2.70	5.20	2.17	-3.85

Number of expansions	4.40	2.88	4.20	2.17	-4.76
Hit-lists examined per iteration	2.25	0.73	1.86	1.00	-21.2
Pool views (Documents judged relevant) per iteration	1.60	1.95	1.60	0.89	-16.4

Table 10 - Search log result averages and standard deviations (topic 442 vs. topic 427). (*percentage of total)

Tables 9 and 10 show the summary information gathered from the search logs, by user group (table 9) and by topic (table 10). In terms of the initial query and search modification the results are varied. There is very little difference in the query size between the groups and the standard deviation is about the same. There is however a more noticeable difference between the topic query sizes i.e. 2.74 for topic 442 compared with 4.80 for topic 427. No user made any spelling mistake in their queries, therefore the test used here cannot differentiate between the groups – more interaction with search terms would be needed to investigate this aspect. Three users in the dyslexia group did use high frequency stop words in their queries however. In terms of query modification by deleting terms there is a difference in behaviour between the control and dyslexic user groups. If we look at the standard deviations, the variation between the users in the dyslexia group is marked (one user deleted only 1 term overall, another deleted 41.5 terms on average of two search iterations – the latter is something of an outlier). Users in the control group tended to be more consistent in term deletion per search ranging from 2.8 to 16.7. The difference in averages therefore is a little misleading. Evidence from the topic results (table 10) shows that users in the topic 442 tended to delete more terms. In terms of query exhaustivity, dyslexic users appear to be doing better but do so on a more ambiguous topic ('heroic acts') with fewer facets. The evidence on query exhaustivity given here does not yield much in the way of difference between the two groups. In terms of topics however, users searching on topic 427 found it easier to find terms which fitted into identified facets.

Turning to interaction with documents, the dyslexic user group viewed 27.4 documents on average compared with 43.8 for the control group (a difference of 16.4 documents on average). In qualitative terms this is a substantial difference. A number of factors could account for this difference, including the slower reading abilities of users in the dyslexia group (we did not test for this aspect however). Variation between users in the two groups is broadly similar. However there appears to be a topic effect, e.g. 35.8 documents were examined per session for topic 442 compared with 28.2 for topic 427, with a much larger standard deviation for the former (21.8 compared with 10.7). Control users judging behaviour appeared to be different, in that they tended to judge more documents to be relevant to the need than dyslexic users (52% for control users, 46% for dyslexic users). This is in part a factor of viewing the total number of documents for a given topic (the assumption being that the more documents you read, the more likely you are to be familiar with the topic). Some evidence for this is provided when looking at topic 442, where more documents were read the users marking 62% as relevant compared with 31% for topic 427. The consistency on this variable for the two views (group and topic) is marked. There is some evidence, that within an iteration dyslexic users view more for a given search (8.32 compared to 6.42 for control users – just under two documents per iteration). So while the dyslexic users viewed fewer documents overall, they appeared more active within a search iteration in terms of document views. This factor might be connected to the difficulty dyslexic users face when decoding words, but there are implications for search which requires investigation. In that, by taking longer to examine a document and making a relevant judgement on it could mean that a search will be more precise. The effect of this is that the search will be narrowed down thus improving precision, at the cost of recall and finding other types of relevant documents. When looking at the per iteration figure for topic, more documents were examined on

average in topic 442 (7.69 documents) than topic 427 (5.99 documents), suggesting that there may be an important topic effect, providing more evidence of the effect on precision and recall. In terms of changing views on relevance, the control group did not change their minds about relevant documents, while there was very little activity in this regard with respect to the dyslexic group. In terms of topics there was no difference in this variable. When looking at the level of agreement with TREC relevance assessments, dyslexic users had much less agreement than controls (see table 9), but the standard deviation is broadly similar and quite high for both user types. However there is clearly a topic effect in that users searching on topic 442 had much less agreement with TREC assessments which is a less focused topic than 427 – just over a half (see table 10).

Turning to the session data, there was very little difference in session lengths between the two groups – this confirms that we managed to keep the search sessions around the same length for all users (all between the required length of 20-30 minutes). The evidence from the topic analysis (table 10) show that sessions on topic 442 lasted longer, but were still within the experimental timeframe. As would be expected with the evidence document views given above, dyslexic users used many fewer search iterations (searches and expands) than control users. There was very little difference between number of search iterations in topics, indicating that this variable could be one which usefully distinguishes between the two user groups. Therefore, while the session lengths were around the same for both groups, search iterations for the dyslexic user group were longer on average. All control users conducted their searches in one session, while two dyslexic users conducted their searches over two sessions, with one user in the group avoiding the expand function from the interface entirely. There is some evidence of wider variation amongst the dyslexia user group from both the standard deviations on the number of sessions and types of searches that were undertaken. Control users tended to view more hitlists per iteration (2.27 has against 1.84 for the dyslexic user group), and variation within this group was lower. With respect to topics, users tended to view more hitlists in topic 442. Control users also tended to view the documents they had judged relevant more often (two per iteration on average), but the variation is larger on this measure than the dyslexia user group. The figures for topics on this variable are identical which may also be indicative.

In summary, there is some evidence of a difference between the two groups on the session data on the document and session measures. Session length is around the same for both groups, but dyslexic users utilized much less iterations in their searches and each iteration was subsequently longer. Control users view more documents, hitlists and relevance pools overall than their dyslexic user counterparts, but the latter tended to read more documents in a search iteration. The type of search used amongst the control group was more homogenous. The evidence collected on queries is much less conclusive, with little in the way of concrete evidence to support any difference between groups apart from variation in deleting terms – however this may be due to an outlier where one particular dyslexic user removed a lot of terms in their searches. Analysis of the differences between searches on topics, showed some clear topic effects (e.g. facet identification), providing some interesting data on the offset of precision against recall. The analysis also provides an indication of variables (such as number of search iterations) that may be topic neutral and could be used to distinguish between the two user groups.

Post-Search interview

In terms of what users liked about the interface there was no general consensus in either subject groups in this study about what they thought was good about it e.g. the layout of the interface, ability to store information in a session and the extra functionality (such as expand) which provides facilities over and above search interfaces used by the group (e.g. Google). One user did reply that the interface ‘did not annoy them’ in answer to this question. In terms of what the users

disliked about the interface, control users tended to be more specific e.g. “time spend deleting terms”, “deleting terms not always helpful”, “pressing return after entering query”, whereas dyslexic uses tended to be more general in their comments e.g. “limiting” interface, “high learning curve”, “look and feel”. In terms of how the users felt about the interface, both groups were equally positive about it with three users in each group saying it was ‘OK’, compared to two users who stated they “didn’t like it very much”. Comments on what they liked about the interface were more positive for those users who marked it as being ‘OK’ as you would expect with comments such as “easy to work with”, “does the job” and “nice and simple”. Less positive comments included “bland compared with Google” and “inflexibility”. One dyslexic user did not understand the function of the expand operation, a non-standard function found in few interfaces.

In terms of what was missing from the interface and what needed to be improved, controls tended to be more consistent in their outlook. This was largely focused on the ability to manipulate queries e.g. “advanced search”, “manual control over search terms” and “Boolean operators”. All control users wanted more power over their queries, in effect wanting more manual query modification techniques than is currently allowed than the interface. It should be noted that there is strong evidence from previous research on query modification that users “highly discriminatory approach to term selection” when involved in more interactive query expansion “worked against the system’s functionality and severely reduced its effectiveness” (Beaulieu, 1997). Dyslexic users tended to be slightly more interested in the look and feel of the interface wanting more image representation on the interface (possibly due to visual information dominance), highlighting terms in titles and provision of summary of documents in the hitlist. Document clustered by ‘graded’ relevance was also suggested as an improvement, and removing the necessity to press return on query entry (which would necessitate a complete redesign of the interface).

With regard to the topic and how difficult it was to search for information on it, both groups had examples where users either found it difficult or easy to search for their particular topic. All users felt that they managed to find at least a couple of relevant documents to satisfy their information need, apart from one control user (C3) who found problems disambiguating documents relating to ‘UV light’ in relation to ‘eyes’ and to ‘cancer’. This information was closely related to how well the users felt about the system its retrieval of relevant documents – unsurprisingly user C3 felt that the system “did not do a good job of finding documents.”

In summary, there did not appear to be any major difference between the groups. In terms of number both groups were equally positive and negative about the interface. There were slight differences in the comments on the interface and controls tended to be more consistent about what needed to be improved, but there was no overall trend in either group.

Discussion

Before we return to the research question, we recap on the evidence provided so far on the study with respect to the differences that exist between the groups. In terms of qualitative information collected in the questionnaires and interviews there does not appear to be much of a difference between the two groups. All users in the study have a sophisticated understanding of search and evaluating information quality, use the internet a lot and largely use the same types of search (web and online) to fulfil their information needs. With regard to the interface used in the experiment and the topics searched on, there does not appear to be any major difference between the groups apart from a couple of minor comments about the interface and what needed to be improved in order to make searching on it more useful. From the information we have here, there is very little evidence, if any, of a difference between the two groups on this qualitative

information and the impact here does not in our view appear to have played much part in the recorded differences.

The evidence from data collected on query terms in the experiment is not sufficient to support any research question that problems in spelling will affect querying skills, both in terms of the initial query entered and manual query modification. We are therefore unable to answer the research question on query variables. No users made spelling mistakes; therefore no differences between the groups could be measured. There was a difference in the term deletion results, but this appeared to be largely due to an outlier (one dyslexic user deleting a large number of terms in their session). It is clear that a different type of experiment is needed to examine this part of user interaction. This must involve some examination of the topic, as differences are shown in table 10 on above on query and term deletion, with some evidence that the identification of facet elements could also have an impact.

Evidence from the log data showed some evidence of differences in reference to interactions i.e. control users used more search interactions in their sessions, and each search was quicker, and these users viewed more results lists and relevance pools on average. We were able to provide some evidence to help answer the research question. Memory abilities appear to have some impact on searching behaviour. The question here however is what element of dyslexia is causing the difference in interaction with documents – reading abilities or memory differences. Slower reading speeds and short term memory problems may also have an impact on dyslexic users' ability to absorb information from documents in roughly the same session time. It is necessary to conduct a further study in order to disambiguate the effect of the variables.

Conclusions and further work

The evidence from this pilot study suggests that there may be a difference in the information searching behaviour of dyslexic users. The small number of participants and the restriction on the number of topics which can be used for ethical reasons do limit the findings of the study, and the results presented here are by no means conclusive. There is much more work in the area to be done, but this study does provide some evidence on how to move forward in the field, and how to provide the information needed to tackle the research question in more detail More effort is needed to tackle literacy problems on query processing, although there is some research on how to support users in this area (Sitbon et al, 2007). The research presented here does indicate that having dyslexia may effect information searching behaviour. We have more data on eye tracking which is being analysed and may shed some more light on search behaviour – results of this will be reported at a latter date. A further study with more users is necessary to replicate these results, and to explore in more detail which aspects of the underlying cognitive deficits, e.g. slow language processing or poor phonological working memory, are more closely related to differences in IR behaviour.

In summary, from the evidence given above the research needs to be extended in the following areas:

- In order to examine the effect of dyslexia on queries, a different approach is needed as the interface as devised in this study appears to hide users' problems with spelling errors. A revised interface must be used which removes all automatic query expansion facilities and allows the users to modify the query as often as they like and however they like. This would more likely give us information on the effect of language problems on query development.
- We need to collect more detailed information on users prior to the searching experiment which would allow us to better distinguish the variables involved in search interactions,

namely reading abilities, speed of language processing and phonological working memory. This would allow us to better understand the effect of poor short term memory in dyslexic users on information searching, and would allow us to gauge the effect of the other two variables.

- One aspect which needs to be addressed in more detail is the effect of the given topic and the impact this has on the user groups. While only two topics were used for search, there were differences between topics when looking at the different types of variables (e.g. query, document and session). A further experiment is needed which focuses on the given topics and their effect, with more topics, different kinds of topics and more users. Can the topic complexity be matched to severity of dyslexia in terms of retrieval effectiveness?
- One interesting result for these experiments was the apparent difference in relevance assessment behaviour. Why are control users more positive about relevance assessments (52% over total, compared with 46% for dyslexics)? Dyslexic users are reading many fewer documents on average, which may well have an impact on the ability to judge the relevance of documents to a topic. There is some indication of this effect from the topic analysis which showed that users reading more documents judged on one topic also judged more as being relevant. Level of agreement with TREC relevance assessments also indicate a possible topic effect. Are dyslexic users narrowing their search and what impact does this have on precision and recall? What is the impact of this on high recall applications such as law and patent search where users need to be more expansive?

This is by no means an exhaustive list of areas in which dyslexia and IR can be further investigated, but given the results of the experiments here, they give us a good lead into the area. This gives us confidence which tackles our secondary research question – validation of the research methodology proposed. Our long term aim is to better understand how dyslexic people use information retrieval systems and provide methods and tools which assist the information seeking process for dyslexic users. The research we have undertaken here and the further research suggested will hopefully give us a good idea of how to go about doing this.

Acknowledgements

We are grateful to Helen Petrie for her input on the initial part of this research, in particular the design of the pre-search questionnaire. We thank all the participants for their time and effort. The research in this paper has been approved by the City University Senate Ethics committee.

References

Alty, J.L., (2002). Dual Coding Theory and education: some media experiments to examine the effects of different media on learning. In Proceedings of Ed-MEDIA: Denver, Collorado, USA, 42-47.

Al-Wabil, A., Zaphiris, P., Wilson, S. (2008). Examining visual attention of dyslexics on web navigation structures with eye tracking. In proceedings of Innovations in Information Technology Conference, Al-Ain, United Arab Emirates, December 2008. IEEE, 717-721.

Al-Wabil, A. (2009). The Effect of Dyslexia on Web Navigation. PhD Thesis, April, City University London.

Beaulieu, M. (1997). Experiments on interfaces to support query expansion. Journal of Documentation, 53 (1), 8-19.

British Dyslexia Association (2007). Dyslexia research information.
(Available on: http://www.bdadyslexia.org.uk/research.html - visited 22nd June 2009)

British Dyslexia Association (2009). Adult dyslexia checklist.
(Available on: http://www.labda.org.uk/adults.htm - visited 29th June 2009).

DRC: Disability Rights Commission (2004). The Web: Access and Inclusion for Disabled people.
(Available on: http://83.137.212.42/sitearchive/drc/PDF/2.pdf - visited 22nd June 2009).

Gregor, P., Dickinson, A., Maccaffer, A. and Andreason, P. (2003). SeeWord – a personal word processing environment for dyslexic computer users. British Journal of Educational Technology, 34 (3)., 341-355.

Jansen, B.J. Spink, A. Bateman, J, and Saracevic, T. (1998) Real life information retrieval: a study of user queries on the web, SIGIR Forum, 32 (1), 5-17.

Kurniawan, S., and Conroy, G. (2006). Comparing Comprehension speed and accuracy of online information in students with and without Dyslexia. Kurniwan, S. and Zaphiris, P. (eds). Advances in Universal Web Design and Evaluation: Research, Trends and Opportunities, Idea Group publishing, 257-270.

Ling, J., van Schaik, P. (2006). The influence of font type and line length on visual search and information retrieval in web pages. International Journal of Human-Computer Studies, 64, 395-404.

MacFarlane, A., Petrie, H. and Jones, S. (2005). A user study on the effect of Dyslexia on Information Retrieval, to be presented at User Studies workshop, In: A. Bailey, I. Ruthven and L. Azzopardi (eds), Proceedings of COLIS 2005 Workshop on Evaluating User Studies in Information Access, 31-38.

Morgan, E. and Klein, C. (2000). The Dyslexic Adult in a non-dyslexic world, Whurr Publishers.

Silverstein, C., Henzinger, M., Marais and Moricz M. (1999) Analysis of a Very large Web Search Engine Query log., SIGIR Forum, 33 (1), 6-12.

Sitbon, L. Bellot, P., Blache, P. (2007). Phonetic based sentence level rewriting of questions typed by dyslexic spellers in an information retrieval context. In: Proceedings of Interspeech 2007 - Eurospeech, Antwerp, Belgium.

Sackville, A., Blankfield, S. and Davey, J. (2002) Supporting Students with Dyslexia in the Effective Use of C&IT in Their Academic Studies SEDA National Staff Development Conference, November, Birmingham, UK.

Snowling, M.J. (2000). Dyslexia, 2nd Edition. Blackwell Publishing.

Spink, A., Ozmutlu, S, Ozmutlu, H and Jansen, B. (2002). U.S. versus European web searching trends. SIGIR Forum, 36 (2), 32-38.

Vakkari, P., Jones, S., MacFarlane A, and Sormunen E. (2004). Query exhaustivity and Relevance Feedback, *Journal of Documentation*, Vol. 60 (2), 109-127.

Vinegrad. M. (1994). A revised adult dyslexia checklist. Educare no. 48, pp, 21-23.