Looking for genre: the use of structural features during search tasks with Wikipedia.

CLARK, M., RUTHVEN, I., O'BRIAN HOLT, P. and SONG, D.

2012

Looking for Genre: the use of Structural Features During Search Tasks with Wikipedia

[1]Malcolm Clark, [2]Ian Ruthven, [1]Patrik O'Brian Holt, [1]Dawei Song
[1]The IDEAS Research Institute, The Robert Gordon University, Aberdeen, Scotland
[2]Department of Computer Science, University of Strathclyde, Glasgow, Scotland
{m.clark1, p.holt, d.song}@rgu.ac.uk, ian.ruthven@cis.strath.ac.uk

ABSTRACT

This paper reports on our task-based observational, logged, questionnaire study and analysis of ocular behavior pertaining to the interaction of structural features of text in Wikipedia using eye tracking. We set natural and realistic tasks searching Wikipedia online focusing on examining which features and strategies (skimming or scanning) were the most important for the participants to complete their tasks. Our research, carried out on a group of 30 participants, highlighted their interactions with the structural areas within Wikipedia articles, the visual cues and features perceived during the searching of the Wiki text. We collected questionnaire and ocular behavior (fixation metrics) data to highlight the ways in which people view the features in the articles. We found that our participants' extensively interacted with layout features, such as tables, titles, bullet lists, contents lists, information boxes, and references. The eye tracking results showed that participants used the format and layout features and they also highlighted them as important. They were able to navigate to useful information consistently, and they were an effective means of locating relevant information for the completion of their tasks with some success. This work presents results which contribute to the long-term goals of studying the features for genre and theoretical perception research.

Categories and Subject Descriptors

H.3.3. **[Information Systems]**: Information Search and Retrieval, *search process, information filtering.*

General Terms

Measurement, Documentation, Experimentation, Human Factors.

Keywords

Perception, Skimming, Scanning, Text, Genre, Wikipedia, Eye tracking, Features

1. INTRODUCTION

Digital social media on the WWW has rapidly become an exciting method for the communication and rapid exchange of social information and knowledge. Communities of Practice now appear all over the web and amongst the multitudes of collaborative communities Wikipedia has become an interesting and commonly used domain for genre analysis especially in the context of enabling social interactivity and empowering of the online community.

IIIX'12 Nijmegen, The Netherlands.

Literary evolutionary processes in Wikipedia [2] has enabled users to develop new and old variants of standardized information forms, that is, genres. In the context of information interaction and processing, we are using a modern eye tracker to record the ocular behavior and strategies of participants in an academic community to show the ways in which they interact holistically with the layout of the main sections of Wikipedia pages, in multiple forms, during natural and realistic search tasks. In Wikipedia, naturally occurring structures, such as genres, offer rich pickings for participants and Wikis are important tools for researchers in the field of genre because they enable a community of practice to construct textual forms for contextual purposes. We aim to locate a set of features which still belong to the form (or structural) concept and to discover the most important aspects of Wikipedia articles, such as discographies and biographies; lists, lists of lists and so on. New methods for automatic genre retrieval are important to this research. Our intention is not only to look at features of form (layout, such as titles, tables, bullet lists etc.) but also to examine methods of the ways in which participants skim and scan texts. Skimming is a reading method which is used to recognize the main purpose of the text. It is performed at a speed several times faster than conventional reading and is normally used when a reader has a large amount of text to read and does not need to understand every word, for example, when a student has to perform a literature search. This type of technique dovetails neatly with genre and types of theoretical perception (Gestalt, Ecological, i.e. affordances, and Constructivism). These genres are there to reduce cognitive load; there is no need for a person to read an entire text since the genre provides these invariant filtering cues in its structure. However, for this study we shall be exploring **which** features and cues our participants find important during their tasks.

In section 2 we will discuss the main areas of theory which form the foundation of this work: research questions, skimming and scanning and theoretical perception. Section 3 describes our study design which includes the experimental setup, search tasks, procedure, data recorded and design. The results of the eye tracking experiment and analysis are in section 4 and comprehensively illustrates the questionnaire and eye tracking gaze data. Penultimately, in section 5 we describe the summary of results and finish with conclusions and future work in section 6.

2. RESEARCH AREAS

2.1 Introduction

One main contention of this research is that an important aspect of the document structure, that is, the layout or genre, is

understated when considering the skimming and scanning process used by the reader to find out whether the document is relevant.

2.2 Research Questions

1. Where does the participant fixate in the first few seconds of viewing a Wikipedia article? Compare information searching between information box (on right) and contents list (on left).
2. Which structural (invariant) cues/formatting features, if any, do participants identify as being used for completing the information-searching task? For example, titles, summary texts, information boxes, tables, reference lists and contents lists.
3. How 'useful' are whole article classic genres, such as lists, lists of lists and biographies during search tasks?
4. Do/can participants skim or scan particular shapes of features (boxes) of the layout of Wikipedia article texts?
5. Alternatively, do participants 'fixate upon' shapes/features/signs of the layout of Wikipedia articles?
6. Which features are the most used? Main titles, sub-titles, information boxes, lists, references, etc.?

2.3 Skimming and Scanning

According to Rayner and other experts, during a search, looking at a scene or reading, the eyes make many particular movements, such as saccades or fixations. The saccades are defined as the eyes making rapid movements "with velocities as high as 500%" [14]. Regarding fixations, our eyes become still for 200-300 microseconds. This is, of course, a misnomer and slightly misleading because the eyes are never completely still because of a constant tremor, namely nystagmus [14].

The reading of texts, whether these are web pages, Wikis or e-mails, involves many complex processes for a human being, such as those discussed in Section 2.4. Skimming and scanning are regularly employed for the purpose of understanding the substance and form of a text. Skimming is a reading technique that is used to identify the main points in a text in order to arrive at an understanding of the text as quickly as possible [14]. It is performed at a speed many times faster than conventional reading speed and is normally used when a reader has a large amount of text to read within a limited time and does not need to understand every word, for example, when a student has to perform a literature search. An abstract or sub titles, for example, could be skimmed to judge whether a particular article would be useful/relevant for the current research. Scanning is a technique used when looking for something, such as a keyword or phrase, and the participant moves their eyes amongst the text. Scanning is usually employed when looking at words, numbers or letters. For example, a student is looking for a definition that is known to exist in amongst the text of an academic article. Sometimes the formatting of words (italics or emboldened) aids the reader to easily identify what is being scanned for. But how are features of layout and formatting used and which do readers use in this environment?

In Wikipedia many readers look only at the information box, summary text, lists, sub titles, references, or maybe only keywords as can be seen in Figures 1, 4 & 5. The layout cues, e.g. sub titles, of a biography are explicit and since a reader examines the results of the search by skimming and scanning the structures in the returned documents, the layout is also a perfect mechanism for finding the appropriate content quickly.

2.4 Genre & Theoretical Perception

Attention is guided by genre information, and the abstract of an academic paper allows a filtering decision to be made on whether the article is relevant or not [19]. This can lead to a potential reduction in the reader's "cognitive load": the reader is given the opportunity to decide that s/he need not read a whole document because the genre provides the invariant cues to its relevance in its structure. Thus, in the opinion of Watt [19], genres behave as "affordances" and in essence can be filtered and categorized by form.

JJ Gibson's affordances are intended to describe how meaning and perception are inter-related: he argues in [6] that instead of perceiving objects (for example, texts) and then adding meaning later, there are visual combinations of invariant and distinctive characteristics of objects which provide cues on how to act and behave in relation to these objects(in this case, structured texts). In the case of genre, these invariant properties or features are primarily layout cues, rather than linguistic cues (but, admittedly, can sometimes be both); they occur in two areas and are referred to in this project as shallow (or surface) features and deep features. Where do these features and cues exist? Frow and Gibson seem to agree when they suggest that the cues and features are located between the reader and the text in the "visual array" (Frow) or in the Ambient Optical Array (Gibson). "Genre is neither a property of (and located 'in') texts, nor a projection of (and located 'in') readers; it exists as a part of the relationship between texts and readers, and it has a systemic existence. It is a shared convention with a social force" [5]. This marks an important overlap between the two scholars.

In addition to the issue of investigating features (or invariants) there is also a case for exploring the possible actions which are afforded to the perceiver of documents; this is one of the main tenets of Gibsonian theory: perception for action [6]. The affordances of genre, in our case, could be defined in terms of drawing the attention of the reader (the perceiver) to salient properties of the Wikipedia text which could trigger a decision that a document is relevant to a participant's successful search.

Toms examined 'Textual Affordances' in her thesis, stating: "aspects or characteristics of an object which makes it obvious how the object can be used." In her study, "…textual affordances are those at the point of user-text interaction in a digital text" [15]. Toms & Campbell [18] suggests that genre is viewed as a shape representing an interface metaphor, in which case the visual cues enable a framework to be loaded (possibly like those frames described by [5; 13]). On the other hand, in their study, Toms and Campbell [17] leaned towards the constructivist (perception for recognition) process, since they aimed to contrast the content (function) and form in order to discover whether readers can perceive and process form on its own or need semantic content to identify it. They also aimed to question whether participants used their previous knowledge to identify a text, such as a web page, or used another technique. Toms and Campbell [17] contended that the 'attributes' of a document's genre enable it to be specifically identified and showed that genre features play a significant role in recognizing documents. They performed experiments using form and function (content), exposing participants, with backgrounds in information technology and an academic environment, to digital and hard copies of Web documents. They suggest that function

(content) is scanned for repetitive patterns (arguably going by the authors' description it could also have been skimmed) and form is possibly scanned for dominant patterns so that possibly two processes are actually on going at the same time. Function (content) provides semantic hints which demonstrate the purpose of the genres, whereas when the document structure was shown, Toms & Campbell [17] stated that: "participants had to match their sensory response with the corresponding representation stored in long-term memory". They also claimed, first of all, that in order to identify a document using form, the reader scans and translates some or all of the visual cues present at the same time to locate the semantic clues. Secondly, the participants constructed or "loaded a set of expectations" which were founded on the available visual clues in the texts. They argued that perception is a top-down process, in contrast to the ecological, bottom-up, where the readers recognize the genres through the attributes of the layout which forms the basis of document recognition (or perception for recognition), and although Toms and Campbell, like Lakoff [11], refer to the bottom-up process and suggest that genres may "act as a single gestalt" [17] they do not explore other possibilities, such as, perception for action and how a genre is perceived when the document is displayed to a reader (in all fairness, it should be pointed out that Watt [19] also fails to explore the perception for recognition concept). In their conclusions, however, Toms and Campbell [17] query how the form of the document affects a reader in the first few seconds of the interaction and this begs the question: how do the form features of a genre aid in text interpretation and use? This is one of the questions that form the central part of our research.

In a later study, Elaine Toms [16] claims that form is important, (but reinforces her 'perceptual' claims) where she explains: "Because the form takes on a distinctive visual appearance, document form essentially represents the shape of a document. That shape is likely two-dimensional since people did not seem to need the multi-dimensional qualities present in the print world to distinguish a shape. Ultimately, the unique shape triggers a user's mental model of that class of genre. In interpreting the shape, a user may develop a set of expectations about the document without first having to read the semantic content".

Although Toms & Campbell [17; 18], Toms [16] and then Watt [19], seem to indicate a leaning towards one perceptual process or another at different times, it may emerge that the processes are both correct, but for different information searching tasks and in different contexts. It is highly likely that documents are identified and used according to differing methods depending on the context of the task, the skill and expertise of the reader, the reading and the use. Such assumptions could be:

- If the reading task is to be performed quickly, skimming is important, but if more time is available, more intensive reading might take place.
- If a participant is looking for a familiar text already seen, then the recognition process (scanning) is important but if the search is a fresh task looking for a particular genre then the ecological process could be vital, to save time.
- It is possible that documents are identified and used according to differing methods depending on the context of the task, the skill and expertise of the reader, the reading and the use.

3. EYETRACKING STUDY

3.1 Experimental Setup

We conducted a task-based observational, logged and questionnaire study using the online version of the English version of Wikipedia as it was in November 2011.

An experimental design was used and 30 participants took part; each was paid £10. The starting point of each task for each participant was the main page of Wikipedia. Participants then had to input an initial search query of their own choice into the search engine provided by Wikipedia. In order to be able to enrich the types of data and the wide range of genres, we decided to use a total of 6 (see 3.3) tasks. The first 15 participants were tested with tasks 1-3, and the subsequent 15 participants were allocated tasks 4-6. Prior to beginning the tasks, each subject was given a three-minute introduction to the eye tracker and a guidance sheet on what was to be expected. Each participant was shown the main page of Wikipedia and the location of the search engine box on the site. Each person was asked to sign a consent sheet before being calibrated to the system. The experimental setup of the evaluation was based on commonly used standards as detailed in previous task-based evaluations, such as [7; 21]. The experimental procedures, such as time given for tasks and questionnaires, were based on the methods and the protocols used in previous interactive experiments [3; 7; 9; 10; 20; 21].

3.2 Our Apparatus

The apparatus used in this study was the T60 model manufactured by Tobii systems. The T60 allows a 60Hz data-sampling rate, which is ample for information seeking studies. The eye tracker is integrated within a 17" TFT monitor, so that intrusion on the participant is negated.

3.3 Search Tasks

3.3.1 Tasks

We constructed six simulated work/situation [8] tasks, in total, that were related to typical tasks to reflect similar participants' needs and were therefore representative of some of the most commonly submitted queries. The tasks were simulated in order to suggest that each participant was preparing to perform an evaluation of end-products task, such as creating an essay, etc. as shown in the examples in [8], to reflect realistic participants' needs. We conducted a small interview on Survey Monkey which was circulated around a football chat forum, around the University of Strathclyde, family and friends by e-mail, and on the Facebook website. There were 53 respondents and they all recalled previous topics and tasks that they had used with Wikipedia. We piloted the indicative request task types [8] and e.g. [4; 20] but this type returned next to no useful data as the tasks were completed too quickly. We decided to use the type of tasks that were demonstrated in [4] using realistic topical tasks supplied by our online survey. In order to get a good range of different types of participants and to avoid any bias in the selection procedure we advertised the experiment throughout the university campus and chose the first 30 people who replied.

To prevent task bias and learning effects, all tasks were allocated randomly by applying the 3x3 Latin square matrix for the first and second group of 15 participants. The 6 tasks are as shown below:

1. You are joining a debating society and need some notes to make a PowerPoint presentation on the first

topic, which is: "Cannabis: Good or Bad?" Since being made illegal in the UK in 1928 and since the introduction of the 1971 Dangerous Drugs Act, the use of cannabis for medicinal reasons has been restricted. However, in recent years, some countries (for example, Austria) have legalized the smoking/ingesting of cannabis by certain patients for pain relief and other medicinal benefits. Thus 'medical cannabis' has become a topic of hot debate. You want to understand the arguments for and against the use of marijuana for medical purposes. Therefore, you decide to do some preliminary research on this subject using Wikipedia. What are the possible health benefits and health problems that may entail from smoking/ingesting cannabis for medical reasons?

2. You have been tasked to write an essay on the Arab Spring which started to be reported in late 2010. The beginning of the so-called 'Arab Spring' led to a huge wave of demonstrations and uprisings in at least 17 countries that has resulted in many long-standing military regimes being overthrown and, in some cases, in civil war. Use Wikipedia to find out some useful information that you feel is appropriate and can be used later to form a basis for the essay. For example, the countries involved and so on.
3. You are in the third year of a social studies degree and have been given coursework on the topic of 'Philanthropy'. On the 4th August 2010, thirty-eight US billionaire philanthropists pledged at least 50% of their wealth to charity through a campaign started by the investor, Warren Buffet, and the Microsoft founder, Bill Gates. Some of those who have signed the pledge include Michael Bloomberg and George Lucas. Many mentioned in 'The Giving Pledge' project are among the most influential people in the contemporary United States and debatably the world. Your coursework states that you have to carry out an investigation to find out who you think is the most influential philanthropist in the pledge group.
4. You are working for ITN news as an intern. There has been a major air crash at an international airport. The news editor wants you to search for background information on the previous top two worst air disasters in history, such as the numbers of fatalities, casualties and so on. She also wants to know the names of airlines with the best and worst safety records.
5. You are on work experience at the sports desk at The Guardian newspaper and have been asked by the editor to collect information on the two rival teams, Boca Juniors and River Plate, as they face each other in the Argentine Cup Final. Use Wikipedia to find out appropriate information about each club, such as the stadiums, star players and the managers of each team.
6. You are in the third year of a political studies' degree course and have been given coursework on studying the legislature in an African country. You decide to focus on Namibia. Collect information about the Parliament, National Council of Namibia, National Assembly and any other information you think is relevant to form the basis of your work.

Each participant was allocated a maximum of 20 minutes for each task, and not one of them exceeded the time cap.

3.3.2 Task statistics

92% of the participants stated that the tasks were clear; 6%, unclear; 2% neutral. 76% of the participants found the tasks easy and 14% had some difficulty. 75% of the participants found the tasks realistic; 12.5%, unrealistic and the reminder were neutral. 73% of the participants were sure they had succeeded in their set tasks, 18% were uncertain and 8% said that they had not succeeded. As regards their understanding of the tasks, 91.5% stated they had understood completely, while 8.5% had not. According to 98% of the participants they enjoyed the experience and this was underlined by the comment: "I found the experimental environment laid back and conducted in a relaxing mode".

3.4 Procedural Task

The study was conducted on a one-to-one basis, but the observer did not intervene unless it was necessary to resolve a problem with the eye tracker.

The procedure was as follows:

1. Entry questionnaire
2. Search tasks (repeated 3 times):
    a. Allocated search task -- Save file(s) to folder or relevant selected text to Word file – Task Questionnaire
3. Exit questionnaire.

They were each told to use only the Wikipedia engine and not the search toolbar in the Internet Explorer (IE) browser. The Tobii system currently restricts the user to using IE8, as it is the only browser object in the SDK.

3.5 Data Recorded

Three main kinds of data were recorded for the experiment - eye gaze data, questionnaires and search task data. We used three kinds of questionnaires: an entry questionnaire, a post-search task (one questionnaire was completed after each task) and an exit questionnaire. The Tobii software records large quantities of types of data, such as logging, gaze plots and heat maps.

1. Logging - including fixations, pupil dilations, queries, mouse clicks, screenshots, video playback, URL, titles of webpage, timestamps, and x/y location of the eye. Saccadic and scan path data is not available at present from the Tobii software automatically (apart from visually in gaze plots and heat maps) as it requires time consuming manual sorting from the logs but will be analyzed in future work.
2. Gaze plots (Figure 1, 4 & 5) - visualize the movement sequence and position of fixations and saccades on the stimulus. The size of the fixation indicates the fixation duration whereas the number on the fixation 'dot' represents the order in which the fixation occurs in the scan path. Gaze plots can be used to illustrate the gaze activity of one or many subjects over the eye tracking session.

By cross-referencing all the data and conducting a comprehensive analysis of the participants' behavior, we were able to test our research questions as listed in Section 1.1. The task data saved by the 30 participants, consisting of whole Wikipedia web pages and text which were copy-pasted within Word files (extracted from articles).

The three types of questionnaires used 5-point Likert scales and were based on templates used previously [12]. The purpose of the entry questionnaire was the recording of demographic information, such as age, web experience and encyclopedia use. The second questionnaire, filled in by the participant after each of the four tasks, was to record the participant experiences, semantic differentials and evaluation of the task. The exit questionnaire was to compare and contrast the four tasks and search completion results that the subjects had just attempted to complete.

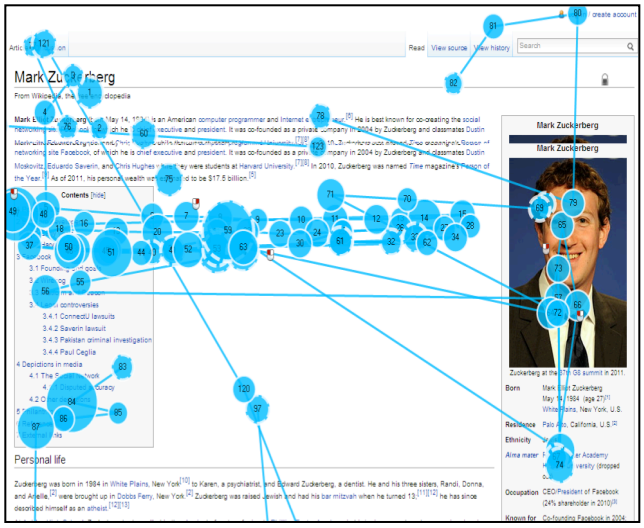

Figure 1. Gaze plots example from Mark Zuckerberg article. The dots are fixations, larger dots mean greater fixation durations and the lines between the fixations are saccades.

3.6 Design

3.6.1 Variables

The examination of differences between genders, ages and nationalities were not really possible for this study as the genders were skewed two-one and ages and nationalities not varied enough. There are two variables in this study. Firstly, the Areas Of Interest (AOI) are the areas on the stimulus which are of most interest to study. Secondly, we require looking at which types of structures are retrieved and used during the tasks carried out by the participants.

1. AOI: (Bullet list, Information box (Figure 3 top right), Contents list (Figure 3, top left), References, Main title, Sub title, Tables, Image captions and Summary text (top of article).
2. Article type: represented in many ways for example, biography (Figures 1, 3 & 5), list (Figure 4), list of lists, discography, football player, country, timeline..

An example of the difference between Footballer biography and biography are shown in Figure 2. A biography in Wikipedia typically centers on the person but can also evolve into what could be argued is a sub-genre. For example, the biography Figures 1 and 2, in the original form, centers on sections that describe the person's life story: birth, early life, later life, wife, siblings and events leading to death. However, on the right of Figure 2, the biography has been modified and oriented toward the profession of the person, in this case a football player. It is still a biography, but is now a biography that gives details of the person's professional life. This can be classed as a sub-genre of a biography; other sections could be added depending on the profession of the person, for example, the article about a football player provides the particular relevant tables and lists, such as teams, player transfers, goals scored, appearances for clubs, caps for the country, etc.

Football Player (Biographical) First name, Surname, Date of Birth, Place of Birth/country, Height, Position, Current Club, Squad No, Youth/Senior Appearances, goals scored (lists)	**Biography**; Born (where, when), Died (where, age, when – if applicable), Spouse, Early Life, Achievements.

Figure 2. Comparison of football player biography with biography found in Wikipedia [2].

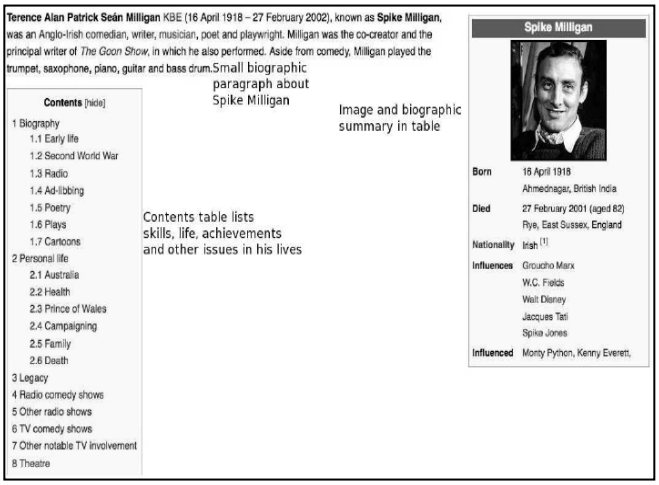

Figure 3. Spike Milligan biography from Wikipedia early 2008 [2]. Summary text at top, information table on the right and contents list on the left.

3.6.2

The interactive measurements are cross-referenced with the research questions (RQ) in section 2.1:

1. Mean fixation count per AOI (RQ 1, 2, 4, 5)
2. Number of times AOI used (RQ1, 2, 4, 5)
3. Total visit (gaze) duration per AOI (RQ1, 2, 4, 5)
4. Visit count (RQ1, 2, 4, 5)
5. Number of articles per task (RQ3)

4. RESULTS & ANALYSIS

4.1 Participants

We recruited 30 subjects and to avoid any bias, we sent out posters and e-mails to the entire university, recruiting the first 30 respondents who replied. The participants were aged between 18 and 42, with a mean age of 23.5. 18 were male and 12, female. 100% of the participants used a computer and the Web every day. They were PhD, post-doc, MSc/MA or undergraduate students working in a variety of fields, such as law, history, computer science and psychology. All participants stated that they used the Web on a daily basis. Roughly 40% of the participants stated that they used online encyclopedias everyday, 31%, once or twice a week and 28%, once or twice a month. Online encyclopedia usage stood at 81% for Wikipedia, and

many other types of encyclopedias, such as Investopedia, Uncyclopedia and Encyclopedia Brittanica etc. were used by a small number of participants, around 3%. Encyclopedic books were only used by 62.5% of participants. Regarding Wikis, 44% of participants stated that they had used Wiki websites in connection with a hobby or coursework. Of the 30 subjects, 87.5% stated that they do not completely trust Wikipedia and, interestingly enough, 78% would not feel comfortable citing it as a source - which means that 9.5% do not trust Wikipedia but may cite it anyway!

4.2 Types of Article

The 30 participants made 396 queries in total (mean of 13.2 each over the whole experiment) altogether over 6 tasks (mean of 2.2 queries per task) and 332 articles in total were retrieved (Table 1). Many errors were made in Wikipedia search engine queries due to mistyping of words, with the result that some of the articles recorded by the eye tracking data had to be discarded because they were Wikipedia error pages, etc. and only 270 could be used for the analysis (126 articles were from direct query results from the Wikipedia whilst the remaining 144 were retrieved through browsing). In addition, some articles did not have any recorded ocular behavior because they were unviewed by the participants. Many other retrieved media were images that were accidentally clicked on by the participant whilst viewing an image, but these will be useful for future research on the ways in which people interact with images on Wikipedia as well as with the small image and captions text. Other discarded contents are listed in Table 1 'non-useful articles'. The breakdown of the entire experiment is shown in Tables 1-3.

Table 1. Articles retrieved in tasks

Articles	Number
Total articles	332
Articles used in analysis	270
Non-useful articles	162
*Wikipedia search results pages	110
*Image pages	18
*Wikipedia query error pages	22
*External web page	1
*Wikimedia	3
*Google search page	1
*Talk page	1

In Table 2 it shows the number of articles retrieved across the six tasks by all participants. The count of relevant and non-relevant article in Table 3 are the total numbers of relevant and non-relevant articles viewed in the study. These counts include duplicates: articles viewed by more than one participant.

Table 2. Number of articles per task

Task	Number
1 (cannabis)	26
2(arab spring)	30
3(philanthropy)	80
4(aircrash)	40
5(football)	53
6(Namibia)	41

Table 3. Articles retrieved in whole study and across all tasks used in analysis

Article Types	Number
Relevant	181 (mean 30.17 SD=1.90 per task)
Non-relevant	89 (mean 15 SD = 1.45 per task)
Biographies (footballers, philanthropists and politicians)	119
Lists	41
Football clubs/stadiums	7
Football stadiums	5
Events	11(air crashes)
Category	5
Timelines (civil war/demonstrations/uprisings e.g. Timeline of 2011 Libyan Civil War)	18
Country/city	7/5
Definition: circa, colo, airline, demonstration, executive, jasmine revolution, judiciary, marijuana. disambiguation: philanthropy, spring)	10
Other misc. articles	42

Before the resulting tables are shown and the results discussed it does of course have to be acknowledged that a higher number of feature fixations may arise due to higher numbers of certain types of features on the pages. For example, longer articles may have a higher number of sub-sections leading to more sub titles.

4.3 Visit Count and Mean Visit Durations per AOI

The AOI visit count (Table 4) starts as soon as a participant first fixates on an AOI, and ends when the participant fixates outside the current AOI. Any number of fixations can occur during a visit. Whenever a participant fixates on text outside the AOI, and then subsequently returns to the AOI, this is added as the beginning of another visit. These are therefore important for our sample, for retrieving relevant information on the tasks. A one-way repeated-measures ANOVA was used to assess mean visit count per AOI and revealed a main effect of $F(1,325) = 265.28$, $p<.001$. Bonferroni post-hoc tests revealed that Sub titles were gazed at most but this could be accounted by feature distribution. Sub titles versus Image captions, Tables, Main titles, References, Numeric lists and Table categories were $p<.001$ apart from Information box $p=.042$ and Bullet list $p<0.05$, but differences between Contents list and Summary text weren't significant. The Contents list was gazed at significantly more than Main title, References, Numeric list and Table categories (all $p<.001$). Summary text was significantly more important than the Main title, References and Table categories (all $p<.001$). The Information box was gazed at less than Sub titles ($p=.042$) and more than Numeric lists ($p=.033$) and Table categories ($p=.006$). The Bullet lists were gazed at less than Sub titles $p<0.05$ and more than Table categories $p=.013$. Image

captions were significantly less gazed at than Sub titles p<.001 but no more significantly gazed at than any other feature. The Tables were less significantly gazed at than Sub titles. The Main title was gazed at less than the Sub titles, Contents list and Summary text (all p<.001).

Table 4. AOI visit (gaze) count and mean visit durations

AOI	Mean visit count	Mean visit durations
Sub titles	63.33	0.74
Contents list	52.09	1.04
Summary text	45.64	0.90
Information box	36.22	1.36
Bullet list	34.48	1.26
Image captions	29.47	0.99
Tables	27.61	1.91
Main titles	14.53	0.42
References	13.48	1.53
Numeric lists	11.57	1.72
Table categories	3.95	0.84

Visit (gaze) duration (Table 4) is the sum of the duration of each fixation within a visit or put simply the duration of each individual visit within the AOI group in seconds. It is occasionally used as a metric of the dissemination of a participant's attention amongst the AOIs. Sometimes this metric is confused due to the number of words in an AOI, as it takes more fixations to process the text. This does not seem to be the case here, since the summary text, which contains the AOI with the largest passages of text in the AOIs only recorded a mean duration of 0.90 seconds which if looked in conjunction with Table 8 ranks fairly low. A one-way repeated-measures ANOVA was used to assess mean visit duration per AOI and revealed a main effect of $F(1,325) = 6.923$, $p<.001$. Bonferroni post-hoc tests revealed that from our sample summarized in Table 4 the participants found the Sub titles more time engaging than the Main title (p<.001) but less than the Table categories (p<.001). The visit durations between the Contents list is more engaging than the Main titles and Table categories (p<.001) but less engaging than the References and Numeric lists (p=.032). The Summary text significantly more engaging than the Main title (p=.005) and Table categories (p=.002). The Information box was fixated upon longer than the Contents list, Summary text and the Main title p<.001. Bullet lists were significantly looked at more than the Main title (p=.003) and Table categories (p<.001). References and Numeric lists had longer durations than the Contents list (p=.032). The durations between Table categories were significant (p<.001).

4.4 Total Visit (gaze) Duration and Fixation Count per AOI

The total visit duration (Table 5) is defined in this context as the duration of all visits within an AOI group even when a user has regressed to the AOI._A one-way repeated-measures ANOVA was used to assess mean total gaze duration per AOI and revealed a main effect of $F(1,325) = 265.28$, $p<.001$. Bonferroni post-hoc tests revealed that Sub titles had effect longer more than Main titles, References, Numeric list and Table categories (all p<.001) and longer than Bullet lists and Image caption also (p=.036). Contents lists and Summary text were gazed at longer than Main title, References, Numeric lists and Table categories (p<.001). The Information box was longer important than the Main title (p=.003), References (p<0.05), Numeric list and Table categories (p<.001). In total the Bullet lists were deemed more important than the Main title (p<0.05), References (p=0.38), Main title (p=.010) and the Table categories (p<.001). The Table had more effect than Main title (p=.010), Numeric list (p=.003) and Table categories (p=.003). The AOI Image captions were deemed insignificant compared to Main titles (p=.010), Numeric list and Table categories (p<.001). The Main title had less effect than the Sub titles (p<.001), Contents list (p<.001), Summary text (p=.002), Information box (p=.003), and Tables (p=.010). The References had less effect than Sub titles and Contents list, both p<.001, Summary text (p=.032) and Information box (p<0.05). The Numeric list did not have less effect than the Image captions. However, the Numeric list AOI did have less effect than Sub titles (p<.001), Contents list (p<.001), Summary text (p<.001), Table (p=.003) and Bullet list (p=.038). Table categories were statistically less effectual than Sub titles, Contents list, Information box and Summary text (p<.001). Bullet lists and Tables were also less effectual (p=.028) and (p=.003) respectively.

Table 5. Total gaze duration in seconds and mean fixation counts per AOI

AOI	Total gaze duration (seconds)	Mean fixation counts
Bullet list	43.37	46.81
Summary text	41.17	62.72
References	20.57	31.18
Main titles	6.17	19.53
Sub titles	46.83	71.32
Information box	49.29	61.08
Tables	52.80	44.45
Image captions	29.25	41.84
Numeric lists	19.9	42.07
Contents list	54.42	52.60
Table categories	3.33	11.50

The counts of fixations on a specific AOI is indicative of the noticeability of the area in question and the cognitive activity of a participant in accomplishing the task. The features most and least prominent are shown in Table 5.

A one-way repeated-measures ANOVA was used to assess mean fixation count per AOI and showed a main effect of $F(1,325) = 24.197$, $p<.001$. We conducted Bonferroni post-hoc tests on the Mean fixation count AOIs. The Sub titles were more noticeable than the Main title (p<.001), Table categories (p=.001), Numeric list (p<.001), References (p<.001), Image captions (p=.001), Table (p=.001) and Bullet lists (p=.009). The Contents list is less noticeable than the Main title (p<.001), References (p=.001), Numeric list (p<.001) and Table categories (p<.001). The Summary text is more noticeable than Main titles (p<.001), Table categories (p<.001) and References (p=.014). The Information box is less noticeable than the Table categories only (p<0.05). The Information box is less effective than Sub titles (p=.009) but more than Main title (p=.001), References (p=.014) and Table categories (p<.001). Bullet lists are less noticeable

than Sub titles ($p<.001$) and more noticeable than Main titles ($p=.039$) and Table categories ($p=.002$). Image captions and Tables were statistically more effective than Table categories ($p=.023$) but less than Sub titles ($p=.001$). Main titles were less noticeable than Sub titles, Contents list, Summary text ($p<.001$), Information box ($p=.001$) and Bullet list ($p=0.39$). References were statistically third least effective and Table categories the absolute least as shown in Table 5.

It could be argued that the AOI Summary text mean fixation counts in Table 5 were so prominent due to the amount of text which features in the captions so some participants were reading the text. After careful and painstaking analysis of each article with the summary text AOIs there were only a few occasions where the text was actually heavily fixated on due to reading which was the case for participants 1, 21 and 27 who between them scored a mean of 74.45 fixations. In other words the large mean amount of fixations was not due to extensive reading but scanning over the text to look for relevant information. The length of the scan paths also signifies this across the text. Short scan path saccades indicate fast reading whereas the long scan path saccades show how the participants were possibly skimming for keywords in the summary text.

4.5 Questionnaire Post Comments

Although there was a wide range of comments about whole article searching, the majority (70%) suggested that they started by forming an initial query and then browsing through the article links in the article web. 12 people suggested that they preferred to search for lists or lists of lists to act as a starting point, particularly for tasks 2, 3 and 4. Regarding task 2, 'The Arab Spring' participants searched for a 'list of countries involved' and for task 4 the participants searched for a 'list of air crashes', 'list of worst air crashes' and 'list of best safety records in airlines'. During task 4, participants submitted queries, such as 'List of Philanthropists Giving Pledge'.

4.6 Structural Features Used?

In the exit questionnaire, the participants indicated how useful they found the structural layout of Wikipedia with regard to helping them carry out the tasks. The percentages certainly back up the data in the tables regarding the usefulness of the structural layout in helping the participants complete the tasks (Table 6)

Table 6. Structure useful

Structure Useful	Percent
Completely useful	56.3%
Quite useful	26%
Not useful	17.7%

The comments below were written by the participants in reply to the question: 'Do you have any further comments about the search experience?':

1. "I enjoy the structure of the pages so finding relevant information was easy"
2. "Layout was very useful and helpful"
3. "Use of boxes to highlight key facts was helpful to finding information"
4. "Wikipedia makes searching very easy as the layout of every page is simple to work with and they all have very useful structures. By providing reference/footnote links it makes the site more reliable".

The participants identified the following features as important though some e.g. hyperlinks were not feasibly marked as AOI due to the sheer amount. This is of course possible but only by examining the hundreds of pages of logs, however, this will be future work. The feature number is in brackets:

Sub titles (24), Contents list (21), Links (18), Tables (18), Information box (top right (12)), Whole articles (10), References (9), Main titles (8), Jumping to Paragraphs and sections (8), Indices (6), Bullet list (4), Emboldened text (3), Index (3).

The feature analysis led to some interesting findings, for example, only 4 participants highlighted bullet lists as important but the gaze data suggested otherwise, for example, Table 4 & 5.

5. SUMMARY OF RESULTS

Our summary of results is as follows:

1. **Where does the participant fixate in the first few seconds of viewing a Wikipedia article? Compare information searching between information box (on right) and contents list (on left).** The data indicates that the most important AOI to our participants in the few seconds of exposure to the Wikipedia articles is firstly the Contents list on the left of the article and secondly the Information Box (for example, Figure 5). Thirdly, though the ocular behavior was possibly skewed on occasion with the Summary text data, due to the amount of textual content, i.e. more fixations through reading, it did indicate higher prominence. However, even taking this into account the data reveals that it was still important over the different metrics examined. In those first few seconds, the structural aspects are very important and, as Toms suggests, could act as textual affordances: the unique shapes may trigger the user's mental model and this interpretation of the shape (or frames) might lead the user to develop a set of expectations about the article before he/she reads the semantic content.
2. **Which structural (invariant) cues/formatting features, if any, do participants identify as being used for completing the information-searching task? For example, titles, summary texts, information boxes, tables, reference lists and contents lists.** During the questionnaire sessions the participants identified the Sub titles (24), Tables (18), Contents Lists (21), Information Boxes (12) and Hyperlinks (18) were identified as the most used during the task. During this experiment it was not possible to apply AOIs to every hyperlink in the Wikipedia pages retrieved so we have no ocular data to record this. The data recorded by the eye tracker reinforces the recollection by the participants of the importance of the Sub titles, Tables, Contents Lists and Information Boxes during the search tasks. There are on occasion indications of the participants not suggesting the usefulness of formatting features but the gaze data suggesting otherwise, such as the bullet lists.
3. **How 'useful' are whole article classic genres, such as lists, lists of lists, biographies (Figures 1, 4 and 5) during search tasks?** Following the analysis of the articles searched and saved by the participants on the desktop which were relevant and not relevant during the tasks, the majority of the articles that the participants used were of a Biographical nature (119) and different types of lists (41) which count as a majority of the retrieved Wikipedia pages during the tasks.

4. **Do/can participants skim or scan (2.2) particular shapes of features (boxes) of the layout of Wikipedia article texts?** During analysis of the data searched by the participants there were instances of both ocular behaviors. In regard to skimming the participants preferred this technique during searches amongst very long documents. For example, the article regarding the Arab Spring is long so most users skimmed the pages to get an understanding if it was an article that was needed for the task. Other examples were demonstrated by the skimming from the Contents list, main title and information box which was a common practice whilst interacting with the articles. This caused the participants to 'skim and scroll' down the articles looking for relevant information pertaining to their tasks. Scanning was a more common behavior during this experiment which is exhibited in the article on Mark Zuckerberg (Figure 1) as it was scanned quite extensively by this specific participant looking for any evidence of 'philanthropy'. Long lists in this experiment that were divided by a large amount of Sub titles (Figure 4) were scanned regularly while looking for keywords or phrases to match the task. The findings will be reinforced by a comparison between saccadic and fixation data in the immediate future.

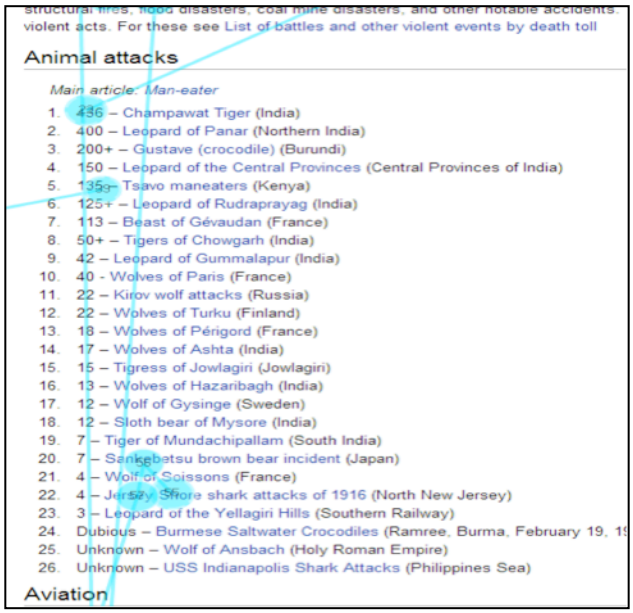

Figure 4. Snippet of article: List of deaths by death toll being skimmed (gaze plots in light blue).

5. **Alternatively, do participants 'fixate upon' shapes/features/signs of the layout of Wikipedia articles?** According to the fixation and gaze data the most commonly visited and fixated areas were the Contents lists, Table categories, References and Information boxes. Again this finding is partially shown in Figure 1 in regard to the Contents list and Information box. The shapes are extremely helpful and natural for the participants to navigate between.
6. **Which features are the most used? Main titles, sub-titles, information boxes, lists, references, etc.?** According to the data the sub titles were most commonly used. In addition the Contents list, Summary text at the top of most articles drew the most attention. The other AOIs were utilized by the participants in certain articles depending on the tasks in preference to other articles, but this would require a much deeper evaluation that is mentioned in 6.1.

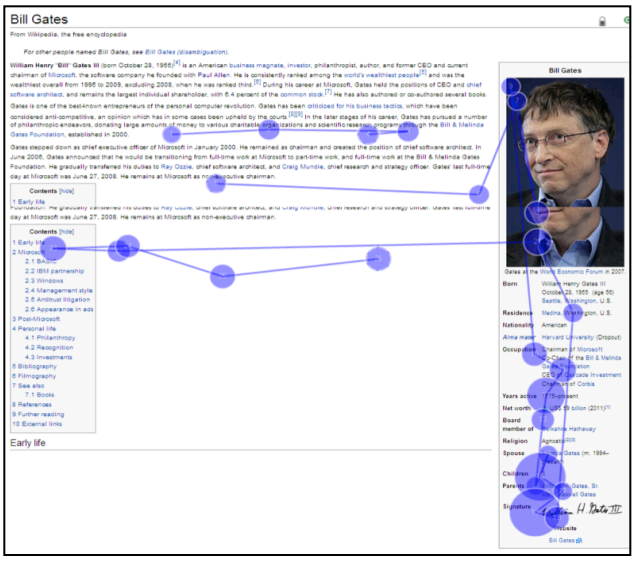

Figure 5. Bill gates biographical article being scanned (gaze plots in purple).

6. FUTURE WORK & CONCLUSIONS

6.1 Future Work

There is much further work that can be explored from the data that was collected from the participants' questionnaires and ocular behavior. For example, we could examine per task and participant analyses, and deeper analysis of genres in terms of large amount of Biographical articles currently retrieved. In terms of Genre does form affect the efficiency with which participants can complete their tasks, in the case of poorly formed documents or do participants use lists to navigate to information? For example, do they navigate lists or use a direct query? A comparison between plain text and the emboldened text for every retrieved article would also be extremely useful to judge between the percentages of formatting used.

A further in depth analysis of lower level engrained features would be appropriate to build on our current findings. For example, on the information boxes are the emboldened titles used more than the actual content, or the bullet symbols used? In an A-Z category article how much attention, if any, do participants pay to the alphabetized label above each section of text the label sits above? Hyperlinks, click-through data, saccadic and scan paths would also be interesting to look at in more depth.

6.2 Conclusions

The experiment described in this paper is the beginning of a larger body of work evaluating how text structure is interacted within and can be used as a means to aiding the user during a search task to find and extract the correct information. We have started to reinforce the work on layout and formatting effects studied in [1; 17; 19] but more analysis and evaluation is needed. Furthermore, we intend to investigate how people use layout in terms of genre and the types of theoretical perception.

However, there is evidence which demonstrates how the features, for example, the shapes of the Information boxes and semantic content are useful as 'textual affordances' to show how the object can be used. Additionally it can also be argued that within Wikipedia when viewing a shape in a document a user constructs a set of expectations then recognizes the validity of the text in the search task. These are just two of the many possibilities which will be examined in the future data analysis. With the aid of further data analysis we can gather clues as to the processes that are used through the recorded ocular behavior, hinting as to the amount of cognitive processing (fixations), or suppressions (saccades). We argue that the layout and invariant cues of the structured texts serve as indicators to direct the reader ultimately towards the task relevant information and can benefit Information Retrieval.

At this stage in our experimental evaluation we cannot yet establish robust conclusions on Genre itself but our results do give strong indications that different features of shapes and textual formatting are being used, and these are important to signify differences in Genre (as discussed in previous studies in section 2.4) so need to be investigated further. We believe this can be achieved by evaluating the data by many other ways, such as task-by-task, biographical, lists and by also evaluating our recently collected web page data alongside this data that has been collected for this study.

7. ACKNOWLEDGMENTS

This research is part of the AutoAdapt research project. AutoAdapt is funded by EPSRC grants EP/F035357/1 and EP/F035705/1. Huge thanks to Ann Mackay for tireless proof reading and we thank the anonymous reviewers for valuable feedback.

8. REFERENCES

[1] Clark, M.J., Ruthven, I., and Holt, P., 2010. Perceiving and using genre by form – an eye-tracking study. *Libri: International Journal of Libraries and Information Services* 60, 3 (September 2010), 268-280.

[2] Clark, M.J., Ruthven, I., and Holt, P.O.B., 2009. The evolution of genre in wikipedia. *Journal for Language Technology and Computational Linguistics* 25, 1, 1-22.

[3] Dupont, G., Requier, S.A., Adam, S., Lecourtier, Y., Grilheres, B., and Brunessaux, S., 2010. A step toward an adaptive composition of query suggestion approaches. In *Proceedings of the 3rd Symposium on Information Interaction in Context* (New Jersey 2010), ACM, 271-276.

[4] Elsweiler, D. and Ruthven, I., 2007. Towards task-based personal information management evaluations. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Amsterdam, The Netherlands 2007), ACM, 23-30.

[5] Frow, J., 2006. *Genre*. Taylor & Francis, New York.

[6] Gibson, J.J., 1986. *The ecological approach to visual perception*. LEA, New Jersey.

[7] Harper, D. and Kelly, D., 2006. Contextual relevance feedback. In *Proceedings of the 1st Symposium on Information Interaction in Context* (Copenhagen, Denmark 2006), ACM, 129-137.

[8] Kelly, D., 2009. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval* 3, 1-2, 224.

[9] Kelly, D., Harper, D., and Landau, B., 2008. Questionnaire mode effects in interactive information retrieval experiments. *Information Processing & Management* 44, 1, 122-141.

[10] Kelly, D., Wacholder, N., Rittman, R., Sun, Y., Kantor, P., Small, S., and Strzalkowski, T., 2007. Using interview data to identify evaluation criteria for interactive, analytical question-answering systems. *Journal of the American Society for Information Science and Technology* 58, 7, 1032-1043.

[11] Lakoff, G., 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press, Chicago, US.

[12] Liu, H., Mulholland, P., Song, D., Uren, V., and Rüger, S., 2010. Applying information foraging theory to understand user interaction with content-based image retrieval. In *Proceedings of the 3rd Symposium on Information Interaction in Context* (New York, USA 2010), ACM, 135-144.

[13] Paltridge, B., 1997. *Genre, frames and writing in research settings*. John Benjamins Publishing Co., Amsterdam.

[14] Rayner, K., 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124, 3, 372-422.

[15] Toms, E.G. 1997 *Browsing digital information examining the affordances in the interaction of user and text* [Doctoral Thesis]. University of Western Ontario.

[16] Toms, E.G., 2001. Recognizing digital genre. *Bulletin of the American Society for Information Science and Technology* 27, 2, 20-22.

[17] Toms, E.G. and Campbell, D.G., 1999. Genre as interface metaphor: Exploiting form and function in digital environments. In *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences* (Hawaii, US, 05 January 1999), IEEE Computer Society, 2008-2024.

[18] Toms, E.G. and Campbell, D.G., 1999. Utilizing information "shape" as an interface metaphor based on genre. In *Proceedings of the 27th Annual Conference of the Canadian Association for Information Science* (Quebec1999), QB: The CAIS, 370-386.

[19] Watt, S.N.K., 2009. Text categorisation and genre in information retrieval. In *Information retrieval: Searching in the 21st century*, A. Göker and J. Davies Eds. John Wiley & Sons, Chichester, U.K., 159-176.

[20] White, R., Ruthven, I., and Jose, J., 2002. The use of implicit evidence for relevance feedback in web retrieval. In *Proceedings of the 24th BCS-IRSG European Colloquium on IR Research: Advances in Information Retrieval* (Glasgow, Scotland 2002), Springer-Verlag, 449-479.

[21] White, R., Jose, J., and Ruthven, I., 2006. An implicit feedback approach for interactive information retrieval. *Information Processing and Management* 42, 1, 166-190.