TKK Reports in Information and Computer Science
Espoo 2008 TKK-ICS-R10

REPORT ON FORMS OF ENRICHED RELEVANCE FEEDBACK

Deliverable D1.1 of FP7 project n° 216529 PinView

He Zhang, Markus Koskela and Jorma Laaksonen

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

TKK Reports in Information and Computer Science
Espoo 2008 TKK-ICS-R10

REPORT ON FORMS OF ENRICHED RELEVANCE FEEDBACK

Deliverable D1.1 of FP7 project n° 216529 PinView

He Zhang, Markus Koskela and Jorma Laaksonen

Helsinki University of Technology
Faculty of Information and Natural Sciences
Department of Information and Computer Science

Teknillinen korkeakoulu
Informaatio- ja luonnontieteiden tiedekunta
Tietojenkäsittelytieteen laitos

Distribution:
Helsinki University of Technology
Faculty of Information and Natural Sciences
Department of Information and Computer Science
P.O.Box 5400
FI-02015 TKK
FINLAND
URL: http://ics.tkk.fi
Tel. +358 9 451 1
Fax +358 9 451 3369
E-mail: series@ics.tkk.fi

ISBN 978-951-22-9669-9 (Print)
ISBN 978-951-22-9670-5 (Online)
ISSN 1797-5034 (Print)
ISSN 1797-5042 (Online)
URL: http://lib.tkk.fi/Reports/2008/isbn9789512296705.pdf

TKK ICS
Espoo 2008

ABSTRACT: This report presents a literature survey conducted to review the current state of the art in research concerning the use of *eye movement measurements* and other non-conventional and *implicit relevance feedback* modalities in content-based image and information retrieval. We define and elaborate on the concept of *enriched relevance feedback* and study its applicability in the intersection of the aforementioned research areas.

KEYWORDS: enriched relevance feedback, implicit relevance feedback, eye movement measurements, content-based image retrieval

ACKNOWLEDGEMENT: The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement n° 216529, Personal Information Navigator Adapting Through Viewing, PinView. All deliverables of the PinView project are available at `http://www.pinview.eu/`.

CONTENTS

1 OVERVIEW

This is the first Deliverable of the *Personal Information Navigator Adapting Through Viewing*, PinView, project, funded by the European Community's Seventh Framework Programme under Grant Agreement n° 216529. The report constitutes the output of Task 1.1 *Study of different forms of enriched feedback* and aims at widening and deepening the literature survey presented already in the PinView project proposal of May 2007 and the later Annex I, "Description of Work" of the Grant Agreement.

The expertise and preexisting experiences of all PinView partners in the topics of content-based image retrieval, relevance feedback and eye movement measurements have been gathered in this report. As such it will serve for the project partners and the public audience as an update of the current state of the art in the research field as well as a collection of and introduction to the essential open scientific literature.

A central concept in the PinView project is the notion of *enriched relevance feedback* in content-based image retrieval. The present work first defines that concept and then lists and exemplifies its uses in existing information retrieval systems. Finally, it is studied how the enriched relevance feedback information can be stored and transferred in actual implementations of content-based retrieval.

The special focus of the survey is on spontaneous or *implicit* relevance feedback, in contrast with the intentional or *explicit* relevance feedback more commonly employed in human–computer interfaces. Implicit relevance feedback is known to be more difficult to model than explicit feedback because it exhibits greater inter-subject variability.

The objective of the PinView project is in the development of advanced techniques for the analysis of eye movement measurements. Therefore, implicit relevance feedback from gaze patterns are studied in this report most extensively. The PinView project partners already have prior experience in using eye movements for extracting implicit relevance feedback in the setting of text reading.

The work presented in this report will be continued mainly in two PinView tasks. The Task 1.3 *Definition of transport protocol for enriched feedback* will focus on technical aspects needed for implementing the interactive and proactive *Personal Information Navigator*. However, the transfer protocol to be developed will be widely applicable in other use scenarios of implicit and enriched relevance feedback as well. The Task 8.1 *Definition of interfaces for information exchange* in turn will facilitate the integration of the algorithms and software modules that either already exist or will be developed in the PinView project.

2 INTRODUCTION

Relevance feedback (RF) has been utilised in information retrieval (IR) for several decades [32, 31, 1]. Its main concept is that a search engine presents the user a set of search results and the user then somehow assesses their relevance to the search topic. This relevance information is then sent back to the search engine to produce a new query or expand the original one. Often, people need to attentively indicate or answer whether or not the retrieved information is relevant, and thus give *explicit relevance feedback*. With large databases and long retrieval sessions, this will inevitably become a laborious task.

The interest for using *implicit relevance feedback* [15], although less accurate than explicit, has increased in recent years. By using implicit relevance feedback an information retrieval system can unobtrusively record the user's behaviour, such as gaze direction, facial expressions and gestures, and use this information to infer his search preferences. Moreover, a combination of explicit and implicit feedback can even better model the user's potential interests [10].

In this report, our main goal is to find out what kinds of feedback forms have been studied by other researchers and used in proactive information retrieval systems. In the *Personal Information Navigator Adapting Through Viewing*, PinView, project[1] and this report, a special emphasis will be given to the use of eye movements in relevance feedback. This is encouraged by the fact that eye movements have proven to provide useful implicit information on the relevance of displayed text and they can be easily and continuously recorded with modern eye trackers without disrupting the user's attention.

This report is organised as follows. The next section gives a general introduction to content-based image and information retrieval. In Section 4, we first introduce the existing forms of relevance feedback in content-based retrieval and then define the concepts of *implicit relevance feedback* and *enriched relevance feedback*. In Section 5, various kinds of existing enriched relevance feedback forms are studied, focusing on eye movements used as the primary implicit feedback source. In Section 6, data schemes that can be used to store and transfer this enriched information are addressed. Finally in Section 7, we present conclusions and a discussion on the future work in the PinView project.

3 CONTENT-BASED IMAGE AND INFORMATION RETRIEVAL

Content-based image retrieval (CBIR) [34] addresses the problem of finding images relevant to the users' information needs from image databases, based principally on *low-level visual features* for which automatic extraction methods are available. Due to the inherently weak connection between the low-level visual features that the computer is relying upon and the *high-level semantic concepts* that humans naturally associate with images, the task of developing these kinds of systems is very challenging.

Unfortunately, very few assumptions about image content can be made in the case of general images, and the generic low-level features used in CBIR are insufficient or impractical to discriminate these kinds of images well on a conceptual level. This creates a quintessential problem in CBIR, namely the *semantic gap* between the high-level semantic concepts used by humans to understand image content and the low-level visual features used by a computer to index the images in a database. Due to the immense need for effective image retrieval applications, a considerable

[1] `http://www.pinview.eu/`

amount of research has been directed on ways to bridge or at least narrow the semantic gap.

The construction of a CBIR index begins with the extraction of suitable features from the images in the database. A feature refers to any characteristic which, in some way, describes the content of an image. In a broad sense, this includes visual features extracted directly from the raw image data, textual keywords, captions, and annotations, and also other kinds of textual or numeric metadata associated with the image.

The simplest visual image features are directly based on the pixel values of the image. These kinds of features are, however, very sensitive to noise and varying imaging conditions and not invariant e.g. to affine transformations. Visual features of more practical use can be obtained by computing certain characteristics or signatures from the images by using suitable image processing or computer vision techniques. This way, the original dimensionality of the image data is reduced during the feature extraction process. As in dimensionality reduction in general, a good feature maintains those characteristics of the original data which preserve the discriminating power while excluding any redundant information.

Visual features can be extracted either with automatic or semi-automatic methods. Fully automatic feature extraction is appealing for obvious reasons, especially with large or dynamic databases. However, the current level of knowledge in image analysis and pattern recognition techniques is limited and the automatic methods at our disposal cannot always provide sufficient discriminating power for effective image retrieval.

Semi-automatic feature extraction methods, on the other hand, rely on human assistance in tasks like image segmentation. For example, since the recognition of objects in general images is a very difficult task for a computer, manually indicated object contours can be used to enhance shape detection and thus shape-based image indexing. Using semi-automatic methods can lead to notable performance improvements, but—depending on the application—the requirement of human effort can be intolerable. However, we believe that incorporating unintentional or implicit search cues from the user can be used as a practical and convenient method for bringing human intelligence and image recognition skills in the feature extraction process.

4 FORMS OF RELEVANCE FEEDBACK IN CONTENT-BASED RETRIEVAL

A popular method to improve the accuracy of content-based image retrieval by narrowing the semantic gap between the high-level semantic concepts and low-level visual features is to shift from single-round queries to navigational queries. In such a setting, a single retrieval session consists of multiple rounds of user–system interaction and query reformulation. This kind of relevance feedback operation can be considered as supervised learning to adjust the subsequent retrieval process by using information gathered from the user's feedback.

Text-based information retrieval has been intensively studied for decades and the usefulness of relevance feedback has long been recognised in the research field. Therefore, a natural basis for developing relevance feedback techniques for content-based retrieval of other information modalities, such as images and videos, is provided by the methodology of traditional text-based information retrieval.

It can be argued that relevance feedback will be even more suitable for content-based retrieval of non-textual information. There are two main reasons for this. First, more ambiguity arises in interpreting images than text, making user interaction more unavoidable. Second, manual modification of the initial query formulation is much more difficult in general content-based retrieval than with textual queries. Still, the research on relevance feedback in the content-based retrieval setting can be seen as a direct descendant of general interaction research in text-based information retrieval.

In the following, we divide the different forms of relevance feedback into two categories. *Conventional relevance feedback* refers to various forms of feedback obtained from interaction with standard user interface components. These include explicit feedback such as check-box selections, mouse clicks and typed query terms. *Enriched relevance feedback* contains more implicit sources of feedback information from human–computer interaction, such as mouse and cursor movements and interaction timing. Furthermore, additional biometric and non-biometric information about the user, obtained using specialised monitoring hardware, can be recorded and used as enriched relevance feedback.

4.1 Content-based retrieval without relevance feedback

On some occasions, image databases have associated captions or other text describing the image content and these annotations can be used to implement image search by textual queries. A successful example of this approach is the Google Image Search[2]. However, accurate manual annotation of large media databases takes a lot of effort and raises the possibility of different interpretations of the image content. Causes for the different interpretations can be traced back to linguistic, cultural and inter-personal differences between the annotators. In addition, the original purpose for the annotations will affect their usefulness in other search contexts.

As an alternative approach, content-based image retrieval has received considerable research and commercial interest in recent years. The first notable CBIR systems include IBM's *QBIC* [8] and *Photobook* [23] developed at MIT. The standard approach to formulate queries in CBIR is *query by pictorial example*, where the image query is based on an example or reference image from the database itself or it can be provided externally. The task of the retrieval system is then to return images as similar to the example image as possible.

The early CBIR systems mentioned above did not contain any real relevance feedback mechanism. The users were merely given the opportunity to select one of the presented images either as the final outcome of the search or as a new pictorial example for continuing the search. For implementing a genuine relevance feedback mechanism, the users should have the possibility to select more than one relevant images at any time. The system would then accumulate the set of the most relevant images found thus far during the search session. Consequently, the search engine should contain a memory trace and state for each individual query session in order to prevent the same images from appearing more than once.

4.2 Conventional relevance feedback

In the PinView project we are mostly interested in content-based image retrieval that can be implemented in a client–server architecture by using the standard HyperText Transfer Protocol (HTTP)[3]. The user interface should mostly be implementable with the standard HyperText Markup Language (HTML)[4].

The HTML standard defines *forms* as the primary mechanism for collecting and processing user interaction more complex than simple hypertext link activations.

[2] `http://images.google.com/`
[3] `http://www.w3.org/Protocols/`
[4] `http://www.w3.org/html/`

The HTML forms can include *textual input fields* and other controls such as *buttons*, *check-boxes* and *menu choices*. The selections made in a form can be submitted with the HTTP protocol back to the search engine by clicking a *submit* button or similar user interface element. The relevance feedback mechanism provided by the standard HTTP/HTML setting is thus mostly limited to simple "key equals value" pairs, where the values are typically Boolean, either *true* or *false*. We call this starting point as *conventional relevance feedback*. Figure 1 displays a snippet of HTML code where a form and some of its controls are defined.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
   "http://www.w3.org/TR/html4/strict.dtd">
<HTML>
  ...
  <BODY>
    ...
    <FORM method="POST" action="query.cgi"
          enctype="multipart/form-data">
      ...
      <INPUT type="CHECKBOX" name="img123456"/>
      <IMG src="img123456.png"/>
      ...
      <INPUT type="SUBMIT" value="Continue query"/>
    </FORM>
    ...
  </BODY>
</HTML>
```

Figure 1: *A snippet of HTML code implementing a form with a check-box.*

The PicSOM CBIR system[5] developed at TKK is a typical example of a CBIR system where conventional Boolean-valued relevance feedback has been used [17]. Figure 2 shows how the system presents the user with a set of images for relevance assessment. Each image has an associated check-box that can be selected to indicate the positive relevance of that image for the current query task. This is a considerable improvement over the older CBIR systems where relevance feedback was not available at all. However, many experiments have revealed the limits of check-box-based explicit relevance feedback in content-based image retrieval.

4.3 Implicit relevance feedback

The notion of *implicit relevance feedback* has already been used many times in this text. The distinction between explicit and implicit relevance feedback is in some occasions quite clear considering the intentionality of the user's behaviour. For example, when the user marks some images as relevant in the user interface depicted in Figure 2, that is obviously explicit feedback. However, if the human–computer interface is able to record the unintentional order of the consecutive check-box clicks, their timing and the path the mouse pointer traverses between the clicks, then this additional data can be regarded as implicit relevance feedback information.

Human–computer interfaces that make use of eye movement measurements have been studied for long. In most of the studies and implemented systems the eye movements have been used as an explicit and intentional control or input modality to the

[5] `http://www.cis.hut.fi/picsom/`

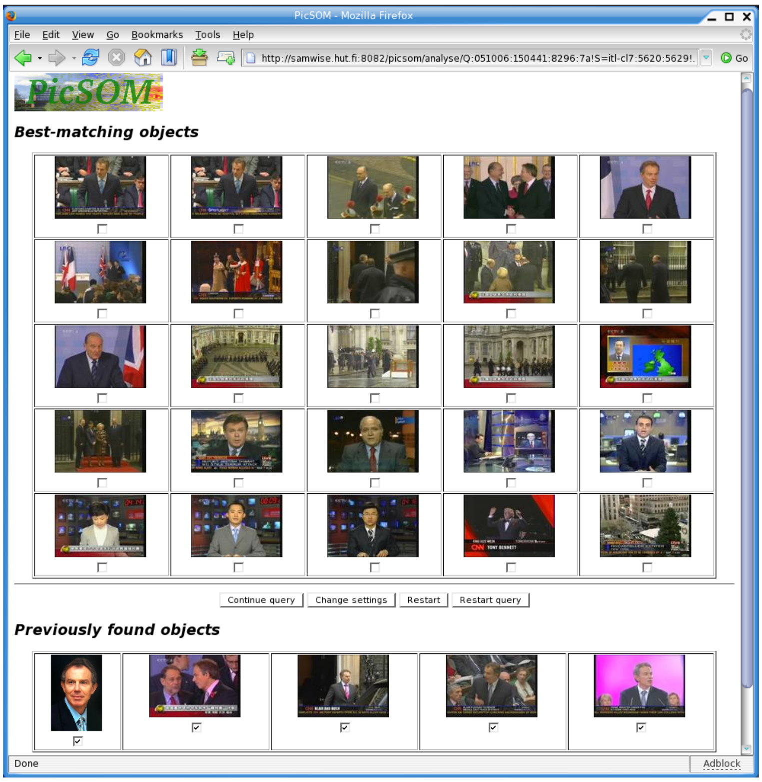

Figure 2: *PicSOM CBIR system utilises conventional relevance feedback with HTTP/HTML check-box forms.*

computer system. In many experiments the gaze-driven input has been found unappealing as it requires persistent concentration in the explicit control task. The eye movements can, however, be used also for implicit relevance feedback where—in the optimal case—the user would not need to be consciously aware of the measurements. It would then be the task of the search engine to extract the relevance information from the unintentional gaze patterns.

4.4 Enriched relevance feedback

By the concept of *enriched relevance feedback* we denote all improvements made to the conventional setting of Boolean-valued relevance feedback implemented with basic HTTP/HTML forms. Instead of just Boolean-valued relevance feedback, the enriched relevance feedback extends the concept of feedback in information retrieval by allowing and merging new feedback modalities from both implicit and explicit sources.

The availability of enriched relevance feedback mechanisms gives more possibilities for inferring the user's potential interests by observing his or her unintentional behaviour such as eye movements. This enriched information can then be processed either on the client side or transferred to the information server for further processing. By using advanced machine learning algorithms, more precise and convenient content-based retrieval can be achieved.

The different combinations of explicit versus implicit and conventional versus enriched forms of relevance feedback are displayed, with some example data sources, in Figure 3. As can be seen, enriched relevance feedback, e.g., eye movements, speech and gestures, can fall into both the explicit and implicit categories. The key difference in those cases is in the intentionality versus non-intentionality of the behaviour.

	Explicit RF	Implicit RF
Conventional RF	Conventional HTTP/HTML: − click−through data − click−location data − text input − radio buttons − check boxes	empty
Enriched RF	Intentional: − eye movements − gestures − speech	Unintentional: − eye movements − gestures − speech Heart rate, blood pressure Body temperature and movements Saving/printing/browsing data

Figure 3: *Classification of various forms of relevance feedback as conventional versus enriched and explicit versus implicit.*

As indicated in Figure 3, enriched relevance feedback may also include human speech and gesture recognition as exemplified by a study of vision-based hand gesture recognition [36]. The described system requires a camera worn on the user's hat or glasses, pointing down to the hand gesture area for tracking the user's hand movements as input information. Human gestures can, however, also be measured by using motion and acceleration sensors.

In a recent study described in [26], movie events and highlights have been detected from unintentional physiological behaviour of subjects watching the movie. The measurements included the heart rate, galvanic skin response, body temperature and movement of the persons. This kind of data could be used as implicit relevance feedback also in content-based retrieval.

In some applications it is possible to analyse event logs that describe the behavioural patterns of the users of that information processing system. For example, it may be possible to study which documents have been downloaded, edited or printed. This kind of information cannot however be used in on-line search tasks as the volume of such data is generally limited and often comes with a considerable delay.

5 EYE MOVEMENTS AND IMPLICIT FEEDBACK

Eye movements have earlier been found to have a strong correlation with human cognitive processes (for a thorough review, see [25]). Historically, a large body of research on eye movements is related to explicit human-computer interaction and control (e.g., [13, 37, 4]). Using eye movements as an implicit feedback source is a relatively new research area. However, eye movements have already demonstrated potential in inferring people's interests in information retrieval tasks and web browsing.

Gaze direction data can be recorded continuously with modern eye trackers. Figure 4 demonstrates eye tracking in a text reading task by a Tobii eye movement recording system. Figure 5 displays the recorded eye movement patterns in that setting. Figure 6 illustrates a recorded gaze pattern when the user has been searching for animals from thumbnail images presented by a web search engine.

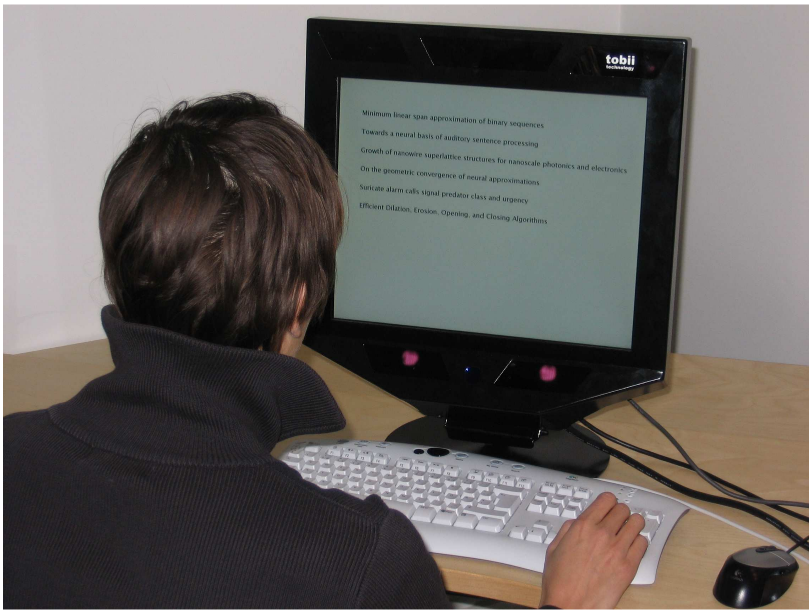

Figure 4: *Gaze tracking in text reading with a Tobii eye movement recording workstation.*

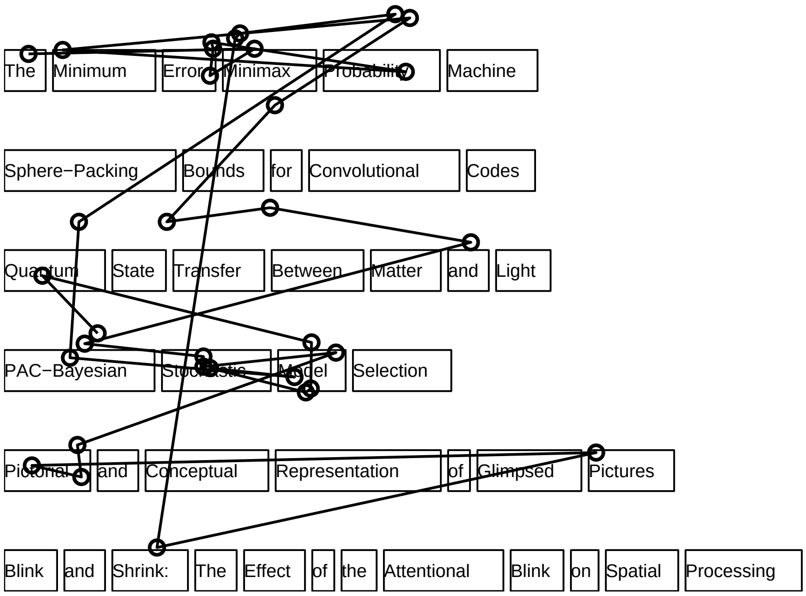

Figure 5: *A recorded eye movement pattern in a text-based information retrieval task.*

In this section, a literature survey is conducted on preexisting works concerning eye movements as an implicit relevance feedback form. In the first subsection we address the question of deriving relevance information from eye movements in proactive information retrieval tasks. In the second subsection we summarise studies on the analysis of user behaviour through eye tracking in a web search scenario.

Figure 6: *A recorded gaze fixation pattern when the user has been searching for animal images.*

5.1 Eye movements used in proactive information retrieval

A pioneer application of using gaze responses to infer the viewer's interests is the *interactive story teller* of 1990 [35]. The story-telling display continuously computes a measure of interest for each part of the image by calculating the number of viewer glances. The objects receiving the highest levels of interest react with a zooming-in and narrated stories.

An elementary relevance judgement is performed in [12] to realize a gaze-assisted translator for reading electronic documents written in a foreign language. When a reader gives more eye focus on some unfamiliar word that he or she encounters, a corresponding translation from an embedded dictionary is triggered.

To our knowledge, the first feasibility study of using eye movements in an information retrieval task was conducted by Salojärvi *et al.* in 2003 [27], where they investigated the possibility to infer from implicit feedback what is relevant for the user's present search topic. In their experimental setting, relevance is controlled by giving the user a specific search task, during which the user's eye movements are measured with an eye tracker. The authors extracted altogether 21 features from raw eye movement signals for each title (sentence) and word, and correlated the features with the known relevance values. Since there is no *a priori* knowledge on which features are the most relevant, the authors then explored the data with statistical information visualisation methods including Principal Component Analysis (PCA) [11], normal Self-Organizing Maps (SOMs) [16] and SOMs that learn metrics [22]. The experimental results show that the relevance of document titles to the retrieval task can be predicted with reasonable accuracy from only a few features (e.g., *fixation count, total fixation duration* and *regression duration*), whereas precise prediction of the relevance of particular words will still require more evolved features and methods.

In a later study by the same authors, Hidden Markov Models (HMM) and Discriminative Hidden Markov Models (DHMM), respectively, are used directly to infer relevance from eye movements for proactive discrimination of relevant texts in [28] and [29]. Both works have a similar experimental settings as that in [27].

In [29], the work has been extended by thorough experiments with a large number of subjects, better equipment that solves the earlier calibration problems, and more detailed analysis of the results. The same data has been used also in the PASCAL Network of Excellence[6] Inferring Relevance from Eye Movements Challenge 2005[7] organised as a competition. The contestants tried to infer the relevance of a read document from the associated eye movement trajectory, aiming at building up a toolbox of robust and efficient methods for relevance extraction [30].

Eye movements can also be combined with other information sources to make a joint prediction of the user's interests. In [24], the implicit feedback from eye movement data is utilised together with a collaborative filtering method to perform proactive information retrieval. In a controlled experiment, 22 subjects were asked to rate their interest in a set of scientific articles. Three of the subjects participated also in an eye movement experiment, in which their eye movements were measured by an eye tracker as an implicit feedback source. Discriminative hidden Markov models were estimated from the recorded data, in which the relevance of the articles was explicitly given by the subjects. Collaborative filtering was then carried out by using the User Rating Profile model [20], a state-of-the-art probabilistic latent variable model computed with Markov Chain Monte Carlo techniques. The experimental results show that for new document titles the prediction accuracy with eye movements, collaborative filtering, and their combination is much better than predictions by chance. Moreover, the combination of eye movements and collaborative filtering produces more accurate relevance predictions than either one technique used alone.

A further information searching strategy based on [24] has been introduced by Hardoon *et al.* [10]. In their work two major tasks have been addressed. One is to construct a query from eye movements alone. With an eye tracker, the eye movements are recorded to formulate an information retrieval query, which is then used to rank unseen documents with respect to their relevance to the current interests of the user. The other task is to construct a query by combining information from implicit relevance feedback from eye movements and explicit relevance feedback. The model they adopt for these two tasks is the Support Vector Machine (SVM) [6], which can be applied to compute reasonable weights in order to predict relevance of unseen documents, and to combine eye movements with textual features in information retrieval. For task one, the performance shows that the average predictions from eye movements alone are better than a random classifier. For task two, a combination of eye movements and textual contents can further improve the precision by about 4%.

In [3], an eye tracker is used to detect read or skimmed document passages. The gaze-based evidence of attention, together with other implicit feedback data such as highlights and comments, is then stored and processed by using Dempster-Shafer theory [33] to derive a uniform degree of attention for any text passage of a document.

In [21], the approach is to use an eye tracker to first identify and extract those keywords that attract relatively longer fixation times from the user. Such keywords are then used for query expansion in textual information retrieval. However, the results show that only a slight benefit in retrieval accuracy can be obtained by taking keywords with long fixation times for query expansion.

[6] `http://www.pascal-network.org/`
[7] `http://www.cis.hut.fi/eyechallenge2005/`

5.2 Eye tracking analysis of user behaviour in web search

The second category of using eye movements considered here is the analysis of user behaviour in a web search task.

A prototype attentive information system [18, 19] named as Simple User Interest Tracker (SUITOR) has been implemented to track user behaviour, model user interests, anticipate user desires and suggest information that might be helpful to the user. The SUITOR gathers information from multiple sources and across multiple modalities ranging from spying of application usage and text typing to tracking of eye gaze and web browsing. These functions are carried out through a set of programs or agents. For instance, the *Investigator* agents can monitor the user's web browsing and eye gaze to determine where on the screen the user is actually reading or just browsing. This information is passed to the *Reflector* agents, which decide what to do about the information discovered by the Investigators and other Reflectors in order to model the user's interests. Finally, the *Actor* agents process information from the Reflectors and perform actions such as displaying information to the user. The SUITOR employs a peripheral display scheme to show the suggestive information considered as relevant, without disrupting the user's current task at hand. Although the performance of the SUITOR system has not been evaluated in the papers, the attentive information system implemented may provide a good framework for the future design of the Personal Information Navigator in the PinView project.

Eye tracking has also been used for the analysis of information usage in web search, which may provide valuable data for designing better web search engines. In [7], a study is presented by using eye tracking techniques to explore the effects of changes in the presentation of search results. The authors find that adding information to a contextual snippet, shown on top of the results, significantly improves the performance in informational tasks, but degrades the performance in navigational tasks. The eye tracking results suggest that when the snippet length is increased, users pay more attention to the snippet and less attention to the URL information located at the bottom of the search result. The authors also suggest improved solutions for designing implementations of search engines. For example, to de-emphasise the snippet one can place the URL below the title, immediately above the snippet or place the snippet in a dedicated pane to the right of the title, URL and other metadata. Furthermore, by using automatic classification, the search engine could classify the search types, whether navigational or informational [7]. In that manner the snippet length and other user interface parameters could be adjusted proactively.

For the purpose of evaluating implicit measures to improve web search, a custom browser [5] has been developed to gather data on implicit interest indicators and to probe for explicit judgements of the visited web pages. The authors find that the amount of scrolling and the time spent on a web page have a strong positive relationship with the user's explicit interests.

Another example has been introduced in [9], where the relationship between implicit and explicit measures of user satisfaction is explored. The data collected with a special browser shows that a combination of the viewing time, click-through data and exiting behaviour of a web page have the strongest correlation with the explicit relevance feedback. These measures are correspondingly the best predictors of users' satisfaction.

The previous two approaches can be reflected with that in [14], where eye tracking data is compared with actual click-through decisions in a web search setting. In that study it was found that the click-through patterns accurately follow the viewing patterns. Consequently, the implicit gaze direction information could to some extent be used as a replacement for link click actions on a web page.

6 STORING AND TRANSFERRING ENRICHED RELEVANCE FEEDBACK

In this section, we briefly discuss the existing schemes we have encountered that can be used for storing and transferring of the enriched relevance feedback data. A general observation has been that in the research papers studied for this literature survey, very little attention has been paid on documenting the used data formats and precise techniques for transferring the data. In the second subsection we then address the data storing and transferring needs in the PinView project.

6.1 Existing schemes

Up to the current time, the simple conventional or non-enriched relevance feedback types, such as click-through or click-location data, have been based on the use of the HTTP transfer protocol and HTML markup language. Unfortunately, the current World Wide Web Consortium (W3C)[8] standards for web pages, XHTML 1.0[9] and HTML 4.01[10], do not provide methods for handling enriched relevance feedback information.

One promising solution for storing enriched relevance feedback data is to use data formats based on the Extensible Markup Language (XML)[11] standard. One practical step towards XML-based eye movement file format standardisation has already been made in the EU FP6 Network of Excellence *Communication by Gaze Interaction*, COGAIN[12]. They have recently developed a common format for carrying eye movement data, see COGAIN's Deliverable 2.2 [2]. In the COGAIN project, the eye movement data is mostly used for explicit control in human–computer interaction. This fact does not, however, cause any major obstacles as the same data format can as well be used for storing implicit relevance feedback information.

In implementing rapid web-based human–computer interaction, a combination of asynchronous JavaScript and the W3C XMLHttpRequest standard[13] has gained prominence. One well-known representative of this approach is the Google Maps service[14]. This scheme provides better usability to web services, for example, a rapid feedback on selection of areas in an image, including highlighting and zooming in. Alternative technologies for implementing rich web-based applications include Adobe Flex[15] and Java. These techniques could likewise be used for transferring the enriched relevance feedback data and new retrieval results between the web-based search engine and its clients equipped with eye trackers.

6.2 Needs in the PinView project

In the PinView project, our aim is to develop a general formalism for storing and transferring enriched relevance feedback information. This will include the definition of the file formats and protocols needed in transferring enriched relevance feedback data from the search clients to the information server. Different data formats and modalities, including still images, video and audio, will be considered in a unifying approach. Most likely the on-line transport will be implemented by using HTTP POST or XMLHttpRequest operations. A strong candidate for the eye move-

[8] `http://www.w3.org/`
[9] `http://www.w3.org/MarkUp/`
[10] `http://www.w3.org/html/`
[11] `http://www.w3.org/XML/`
[12] `http://www.cogain.org/`
[13] `http://www.w3.org/TR/XMLHttpRequest/`
[14] `http://maps.google.com/`
[15] `http://www.adobe.com/go/flex/`

ment and other enriched relevance feedback data format is the COGAIN format.

These questions will become more actual in the PinView Tasks 1.3 *Definition of transport protocol for enriched feedback* and 8.1 *Definition of interfaces for information exchange*, where the practical needs for data storing and transfer in the PinView project are addressed in detail.

7 CONCLUSIONS

Relevance feedback techniques have a long history in information retrieval. The traditional way to give feedback is to respond explicitly whether or not the retrieved information is relevant, which may be quite tedious in practice. Implicit feedback techniques have gained more attention as they can be used to infer the user's potential interests without disruption.

In this report, we have made a new classification of existing relevance feedback techniques into conventional versus enriched forms. A literature survey was conducted on the forms of enriched relevance feedback, in which special focus was given to the user's eye movements utilised as an implicit or unintentional relevance feedback information source. Eye movement data can be obtained both continuously and unconsciously with modern eye trackers.

From the survey we can first conclude that eye movements have shown promising performance as an enriched feedback form for inferring user preferences, both in textual information retrieval settings and in the analysis of user behaviour during web search. Our second conclusion is that most of the research efforts in using eye movements in information retrieval have concentrated in textual search tasks. Very little attention has this far been paid on visual search tasks where, in our opinion, is a clear potential for gaze-based information retrieval.

Our third conclusion is related to the observed lack of existing and widely-adopted standards for storing eye movement data. In the PinView project we will need data formats that are capable for storing also other modalities of enriched relevance feedback. In addition, a transfer protocol for rapidly communicating that information from the client side to the search engine will be required. These practical questions need to be solved before the implementation of the prototype for interactive and proactive Personal Information Navigator can be started.

ACKNOWLEDGEMENTS

We wish to thank Dr. David R. Hardoon of University College London and M.Sc. (Tech.) Mats Sjöberg and M.Sc.(Tech.) Ville Viitaniemi of Helsinki University of Technology for their valuable contributions in the writing of this report and for comments on its draft versions.

References

[1] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley, 1999.

[2] Richard Bates, Howell Istance, and Oleg Spakov. Requirements for the common format of eye movement data. Technical report, Communication by Gaze Interaction (COGAIN): Deliverable 2.2, 2005.

[3] Georg Buscher. Attention-based information retrieval. In *SIGIR Proceedings on Research and development in information retrieval*, pages 918–918, Amsterdam, Netherland, 2007.

[4] Craig A. Chin, Armando Barreto, J. Gualberto Cremades, and Malek Adjouadi. Performance analysis of an integrated eye gaze tracking / electromyogram cursor control system. In *Proc. 9th Int. ACM SIFACCESS Conf. on Computers and accessibility*, pages 233–234, New York, NY, USA, 2007.

[5] Mark Claypool, David Brown, Phong Le, and Makoto Waseda. Inferring user interest. *IEEE Internet Computing*, 5(6):32–39, Nov/Dec 2001.

[6] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge University Press, 2000.

[7] Edward Cutrell and Zhiwei Guan. What are you looking for? an eye-tracking study of information usage in web search. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 407–416, San Jose, California, April-May 2007.

[8] Myron Flickner, Harpreet Sawhney, Wayne Niblack, et al. Query by image and video content: The QBIC system. *IEEE Computer*, 28:23–31, September 1995.

[9] Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan Dumais, and Thomas White. Evaluating implicit measures to improve web search. *ACM Transactions of Information Systems (TOIS)*, 23(2):147–168, 2005.

[10] David R. Hardoon, John Shawe-Taylor, Antti Ajanki, Kai Puolamäki, and Samuel Kaski. Information retrieval by inferring implicit queries from eye movements. In *Eleventh International Conference on Artificial Intelligence and Statistics*, 2007.

[11] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.

[12] Aulikki Hyrskykari, Päivi Majaranta, and Kari-Jouko Räihä. Proactive response to eye movements. In G. W. M. Rauterberg, M. Menozzi, and J. Wesson, editors, *INTERACT'03*. IOS Press, 2003.

[13] Robert J. K. Jacob. Eye movement–based human–computer interaction techniques: Toward non command interfaces. In H.R. Hartson and D. Hix, editors, *Advances in Human–Computer Interaction*, pages 151–190. Ablex Publishing Co., Norwood, N.J., 1993.

[14] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. Accurately interpreting clickthrough data as implicit feedback. In G. Marchionini, A. Moffat, J. Tait, R. Baeza-Yates, and N. Ziviani, editors, *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 154–161, New York, NY, USA, 2005. ACM Press.

[15] Diane Kelly and Jaime Teevan. Implicit feedback for inferring user preference: a bibliography. *SIGIR Forum*, 37(2):18–28, 2003.

[16] Teuvo Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer-Verlag, Berlin, third edition, 2001.

[17] Jorma Laaksonen, Markus Koskela, and Erkki Oja. PicSOM—Self-organizing image retrieval with MPEG-7 content descriptions. *IEEE Transactions on Neural Networks, Special Issue on Intelligent Multimedia Processing*, 13(4):841–853, July 2002.

[18] Paul P. Maglio, Rob Barrett, Christopher S. Campbell, and Ted Selker. Suitor: an attentive information system. In *Proceedings of the 5th international conference on Intelligent user interfaces*, pages 169–176, 2000.

[19] Paul P. Maglio and Christopher S. Campbell. Attentive agents. *Commun. ACM*, 46(3):47–51, 2003.

[20] Benjamin Marlin. Modeling user rating profiles for collaborative filtering. In *Advances in Neural Information Processing Systems*, volume 16, Cambridge, MA, 2004. MIT Press.

[21] Tristan Miller and Stefan Agne. Attention-based information retrieval using eye tracker data. In *Proceedings of K-CAP'05*, pages 209–210, Banff, Alberta, Canada, October 2-5 2005.

[22] Jaakko Peltonen, Arto Klami, and Samuel Kaski. Learning more accurate metrics for self-organizing maps. In *ICANN '02: Proceedings of the International Conference on Artificial Neural Networks*, pages 999–1004, London, UK, 2002. Springer-Verlag.

[23] Alex Pentland, Rosalind W. Picard, and Stan Sclaroff. Photobook: Tools for content-based manipulation of image databases. In *Storage and Retrieval for Image and Video Databases II*, volume 2185 of *Proceedings of SPIE*, pages 34–47, San Jose, CA, USA, 1994.

[24] Kai Puolamäki, Jarkko Salojärvi, Eerika Savia, Jaana Simola, and Samuel Kaski. Combining eye movements and collaborative filtering for proactive information retrieval. In G. Marchionini, A. Moffat, J. Tait, R. Baeza-Yates, and N. Ziviani, editors, *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 146–153. ACM press, New York, NY, USA, 2005.

[25] Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372–422, 1998.

[26] Sandra Rothwell, Bart Lehane, Ching Hau Chan, Alan F. Smeaton, Noel O'Connor, Gareth Jones, and Dermot Diamond. The CDVPlex Biometric Cinema: Sensing physiological responses to emotional stimuli in film. In *Pervasive 2006 - Proceedings of the 4th International Conference on Pervasive Computing*, 2006.

[27] Jarkko Salojärvi, Ilpo Kojo, Jaana Simola, and Samuel Kaski. Can relevance be inferred from eye movements in information retrieval? In *Proceedings of WSOM'03, Workshop on Self-Organizing Maps*, pages 261–266, Kyushu Institute of Technology, Kitakyushu, Japan, 2003.

[28] Jarkko Salojärvi, Kai Puolamäki, and Samuel Kaski. Relevance feedback from eye movements for proactive information retrieval. In J. Heikkilä, M. Pietikäinen, and O. Silvén, editors, *Workshop on Processing Sensory Information for Proactive Systems (PSIPS 2004)*, Oulu, Finland, 2004.

[29] Jarkko Salojärvi, Kai Puolamäki, and Samuel Kaski. Implicit relevance feedback from eye movements. In Wlodislaw Duch, Janusz Kacprzyk, Erkki Oja, and Slawomir Zadrozny, editors, *Artificial Neural Networks: Biological Inspirations – ICANN 2005*, Lecture Notes in Computer Science 3696, pages 513–518, Berlin, Germany, 2005. Springer-Verlag.

[30] Jarkko Salojärvi, Kai Puolamäki, Jaana Simola, Lauri Kovanen, Ilpo Kojo, and Samuel Kaski. Inferring relevance from eye movements: Feature extraction. Helsinki University of Technology, Publications in Computer and Information Science, Report A82, March 2005. The Challenge has a web site at `http://www.cis.hut.fi/ eyechallenge2005/`.

[31] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. Computer Science Series. McGraw-Hill, New York, 1983.

[32] Gerard Salton, editor. *The SMART retrieval system: Experiments in automatic document processing*. Prentice-Hall, 1971.

[33] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton, NJ, Princeton, University Press, 1976.

[34] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.

[35] India Starker and Richard A. Bolt. A gaze–responsive self–disclosing display. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 3–10, Gaithersburg, MD, USA, 1990. ACM Press.

[36] Thad Starner, Joshua Weaver, and Alex Pentland. A wearable computer based American Sign Language recognizer. In *First Internatinal Symposium. on Wearable Computing*, Cambridge, MA, 1997.

[37] David J. Ward and David J.C. MacKay. Fast hands-free writing by gaze direction. *Nature*, 418:838, 2002.

TKK REPORTS IN INFORMATION AND COMPUTER SCIENCE

TKK-ICS-R1	Nikolaj Tatti, Hannes Heikinheimo Decomposable Families of Itemsets. May 2008.
TKK-ICS-R2	Ville Viitaniemi, Jorma Laaksonen Evaluation of Techniques for Image Classification, Object Detection and Object Segmentation. June 2008.
TKK-ICS-R3	Jussi Lahtinen Model Checking Timed Safety Instrumented Systems. June 2008.
TKK-ICS-R4	Jani Lampinen Interface Specification Methods for Software Components. June 2008.
TKK-ICS-R5	Matti Koskimies Applying Model Checking to Analysing Safety Instrumented Systems. June 2008.
TKK-ICS-R6	Alexander Ilin, Tapani Raiko Practical Approaches to Principal Component Analysis in the Presence of Missing Values. June 2008.
TKK-ICS-R7	Kai Puolamäki, Samuel Kaski Bayesian Solutions to the Label Switching Problem. June 2008.
TKK-ICS-R8	Abhishek Tripathi, Arto Klami, Samuel Kaski Using Dependencies to Pair Samples for Multi-View Learning. October 2008.
TKK-ICS-R9	Elia Liitiäinen, Francesco Corona, Amaury Lendasse A Boundary Corrected Expansion of the Moments of Nearest Neighbor Distributions. October 2008.

ISBN 978-951-22-9669-9 (Print)
ISBN 978-951-22-9670-5 (Online)
ISSN 1797-5034 (Print)
ISSN 1797-5042 (Online)