Department of Information and Computer Science

Inference of relevance for proactive information retrieval

Antti Ajanki

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 128/2013

Inference of relevance for proactive information retrieval

Antti Ajanki

Doctoral dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the School of Science for public examination and debate in Auditorium T2 at the Aalto University School of Science (Espoo, Finland) on the 27th of September 2013 at noon.

Aalto University
School of Science
Department of Information and Computer Science

Supervising professor
Prof. Samuel Kaski

Preliminary examiners
Prof. Markku Tukiainen, University of Eastern Finland
Dr. Emilio Parrado-Hernandez, Univ. Carlos III de Madrid, Spain

Opponent
Prof. Kari-Jouko Räihä, University of Tampere

Aalto University publication series
DOCTORAL DISSERTATIONS 128/2013

ISBN 978-952-60-5294-6 (printed)
ISBN 978-952-60-5295-3 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-5295-3

Unigrafia Oy
Helsinki 2013

Finland

441 697
Printed matter

A" Aalto University

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Abstract

Author
Antti Ajanki

Name of the doctoral dissertation
Inference of relevance for proactive information retrieval

Publisher School of Science

Unit Department of Information and Computer Science

Series Aalto University publication series DOCTORAL DISSERTATIONS 128/2013

Field of research Computer and Information Science

Manuscript submitted 17 April 2013 **Date of the defence** 27 September 2013

Permission to publish granted (date) 27 June 2013 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

Search engines have become very important as the amount of digital data has grown dramatically. The most common search interfaces require one to describe an information need using a small number of search terms, but that is not feasible in all situations. Expressing a complex query as precise search terms is often difficult. In the future, better search engines can anticipate user's goals and provide relevant results automatically, without the need to specify search queries in detail.

Machine learning methods are important building blocks in constructing more intelligent search engines. Methods can be trained to predict which documents are relevant for the searcher. The prediction is based on recorded feedback or observations of how the user interacts with the search engine and result documents. If the relevance can be estimated reliably, interesting documents can be retrieved and displayed automatically.

This thesis studies machine learning methods for information retrieval and new kinds of applications enabled by them. The thesis introduces relevance inference methods for estimating query terms from eye movement patterns during reading and for combining relevance feedback given on multiple connected data domains, such as images and their captions. Furthermore, a novel retrieval application for accessing contextually relevant information in the real world surroundings through augmented reality data glasses is presented, and a search interface that provides browsing cues by making potentially relevant items more salient is introduced.

Prototype versions of the proposed methods and applications have been implemented and tested in simulation and user studies. The tests show that these methods often help the searcher to locate the right items faster than traditional keyword search interfaces would.

The experimental results demonstrate that, by developing custom machine learning methods, it is possible to infer intent from feedback and retrieve relevant material proactively. In the future, applications based on similar methods have the potential to make finding relevant information easier in many application areas.

Keywords Machine learning, relevance inference, information retrieval, implicit feedback, eye tracking

ISBN (printed) 978-952-60-5294-6 **ISBN (pdf)** 978-952-60-5295-3

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2013

Pages 180 **urn** http://urn.fi/URN:ISBN:978-952-60-5295-3

Aalto-yliopisto

Tiivistelmä

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tekijä
Antti Ajanki

Väitöskirjan nimi
Relevanssin päättely proaktiivisessa tiedonhaussa

Julkaisija Perustieteiden korkeakoulu

Yksikkö Tietojenkäsittelytieteen laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 128/2013

Tutkimusala Tietojenkäsittelytiede

Käsikirjoituksen pvm 17.04.2013 **Väitöspäivä** 27.09.2013

Julkaisuluvan myöntämispäivä 27.06.2013 **Kieli** Englanti

☐ **Monografia** ☒ **Yhdistelmäväitöskirja (yhteenveto-osa + erillisartikkelit)**

Tiivistelmä

Hakukoneista on tullut erittäin tärkeitä digitaalisen tiedon määrän kasvaessa räjähdysmäisesti. Tavallisin tapa informaation etsimiseen ovat hakusanoihin perustuvat hakukoneet, mutta ne eivät ole hyödyllisiä kaikissa tilanteissa. Monimutkaisen tiedontarpeen pelkistäminen hakusanoiksi on usein vaikeaa. Tulevaisuudessa hakukoneet ovat älykkäämpiä ja osaavat jopa ennakoida käyttäjän tarpeita ja hakea tarpeellisia tietoja automaattisesti ilman, että käyttäjän tarvitsee syöttää tarkkoja hakusanoja.

Koneoppimismenetelmät ovat tärkeitä rakennuspalikoita kehitettäessä älykkäämpiä hakukoneita. Koneoppimisen avulla tietokoneen on mahdollista oppia tunnistamaan mikä tieto on tärkeää. Tunnistaminen perustuu käyttäjän toimien havainnointiin ja käyttäjän antaman palautteen analysointiin. Jos ennustaminen onnistuu hyvin, tietokoneen on mahdollista hakea ja esittää mielenkiintoista tietoa automaattisesti.

Tässä väitöskirjassa tutkitaan tiedonhaussa auttavia koneoppimismenetelmiä ja uudenlaisia sovelluksia, joita menetelmien avulla on mahdollista toteuttaa. Väitöskirjassa esitellään koneoppimismenetelmiä tiedontarpeen päättelemiseen tekstin lukemisen aikana mitattujen silmänliikkeiden perusteella ja menetelmiä, jotka tekevät päätelmiä yhdistämällä eri tietotyypeille kuten kuville ja kuvien otsikoille annettua palautetta. Lisäksi esitellään uudentyyppinen tapa hakea ympäröivään maailmaan liittyvää tietoa datalasien läpi katsottavan lisätyn todellisuuden näkymän kautta. Toinen työssä tutkittava uusi käyttöliittymäperiaate on hakukäyttöliittymä, joka auttaa tiedon selailua korostamalla tärkeäksi pääteltyjä tietoalkioita.

Esitellyistä koneoppimismenetelmistä ja tiedonhakusovelluksista on tehty prototyyppitoteutukset, joita on testattu simulaatio- ja käyttäjäkokeissa. Kokeiden perusteella uudet menetelmät ja sovellukset auttavat usein tiedonhakijaa löytämään halutun tiedon nopeammin kuin perinteiset hakusanoihin perustuvat käyttöliittymät.

Väitöskirjassa esitettyjen tulosten perusteella voidaan todeta, että tiedon tärkeyden päättely ja automaattinen tiedonhaku ovat mahdollisia kehittyneiden koneoppimismenetelmien avustuksella. Tämänkaltaisiin menetelmiin perustuvat hakukoneet voivat tulevaisuudessa helpottaa tiedonhakua monilla sovellusaloilla.

Avainsanat koneoppiminen, relevanssin päättely, tiedonhaku, epäsuora palaute, silmänliikkeiden seuranta

ISBN (painettu) 978-952-60-5294-6 **ISBN (pdf)** 978-952-60-5295-3

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2013

Sivumäärä 180 **urn** http://urn.fi/URN:ISBN:978-952-60-5295-3

Preface

This thesis was done in the Department of Information and Computer Science (former Adaptive Informatics Research Centre) of Aalto University School of Science (former Helsinki University of Technology). I gratefully acknowledge the financial support of the Urban contextual information interfaces with multimodal augmented reality (UI-ART) project of the Aalto MIDE programme; the Helsinki Doctoral Programme in Computer Science – Advanced Computing and Intelligent Systems (Hecse); the graduate school of the Department of Computer Science and Engineering at Helsinki University of Technology; the Pervasive Information, Interfaces and Interaction (PI3) and the Cross-linking Visual Information activities of the EIT ICT Labs KIC; the Pattern Analysis, Statistical Modelling and Computational Learning (PASCAL2) EU Network of Excellence; the Devices and Interoperability Ecosystem (DIEM) programme of the Strategic Centre for Science, Technology and Innovation in the Field of ICT (TIVIT SHOK); and the Nokia Foundation. I have been a member of Helsinki Institute for Information Technology HIIT and the Finnish Centre of Excellence in Computational Inference Research (COIN).

I would like to express my warmest thanks to my instructor and supervisor, Prof. Samuel Kaski. He provided me an opportunity to work on machine learning and proactive information retrieval and taught me an immense amount about science and research.

I am thankful for the pre-examiners of my thesis, Prof. Markku Tukiainen and Dr. Emilio Parrado-Hernandez. Their comments helped me to improve the thesis.

I am indebted to all my co-authors. I am especially grateful for Docent Kai Puolamäki for guiding me at the beginning of my doctoral work, and for Dr. Melih Kandemir, Dr. Markus Koskela and Docent Jorma Laaksonen. It has been a pleasure to work with you. I also thank my M.Sc. thesis

instructor, Docent Janne Nikkilä, for showing how exciting research is.

I would like to thank all past and present members of MI group and the rest of the colleagues at the department for creating an inspiring work environment. I am also thankful for Tarja Pihamaa and Leila Koivisto, who have helped me on administrative issues on numerous occasions.

Finally, I am grateful to my parents for supporting and encouraging me in my pursuit.

Espoo, August 15, 2013,

Antti Ajanki

Contents

List of publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Antti Ajanki, David R. Hardoon, Samuel Kaski, Kai Puolamäki and John Shawe-Taylor. Can eyes reveal interest? – Implicit queries from gaze patterns. *User Modeling and User-Adapted Interaction*, 19(4):307–339, 2009.

II Kai Puolamäki, Antti Ajanki and Samuel Kaski. Learning to learn implicit queries from gaze patterns. In *Proceedings of the 25th International Conference on Machine Learning (ICML)*, pages 760–767, ACM, New York, NY, 2008.

III Antti Ajanki, Mark Billinghurst, Hannes Gamper, Toni Järvenpää, Melih Kandemir, Samuel Kaski, Markus Koskela, Mikko Kurimo, Jorma Laaksonen, Kai Puolamäki, Teemu Ruokolainen and Timo Tossavainen. An augmented reality interface to contextual information. *Virtual Reality*, 15(2-3):161–173, 2011.

IV Antti Ajanki and Samuel Kaski. Probabilistic proactive timeline browser. In *Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN)*, pages 357–364, Springer-Verlag, Berlin, Germany, 2011.

V Antti Ajanki, Markus Koskela, Jorma Laaksonen and Samuel Kaski. Dynamic browser for personal history. Accepted for publication in *15th*

ACM International Conference on Multimodal Interaction (ICMI), 8 pages, 2013.

VI Antti Ajanki, Mehmet Gönen and Samuel Kaski. Multi-domain ranking. Submitted to a journal, 15 pages.

Author's contribution

Publication I: "Can eyes reveal interest? – Implicit queries from gaze patterns"

The author participated in planning the model and the experiments. The author carried out the user experiments. Writing of the paper was a collaborative effort.

Publication II: "Learning to learn implicit queries from gaze patterns"

The author participated in planning the model and experiments, implemented the model and participated in supervising the user experiments. Writing of the paper was a collaborative effort.

Publication III: "An augmented reality interface to contextual information"

The retrieval framework was designed together. The author implemented the retrieval subsystem and had a major role in integration of other subsystems. The author participated in design and execution of experiments. Writing of the paper was a collaborative effort.

Publication IV: "Probabilistic proactive timeline browser"

The author implemented the retrieval system and carried out the experiments. Writing of the paper was a collaborative effort.

Publication V: "Dynamic browser for personal history"

The principles of the dynamic timeline interface and the experiments were designed together. The author had a major role in implementing the prototype system and carried out the user studies. Writing of the paper was a collaborative effort.

Publication VI: "Multi-domain ranking"

The multi-domain ranking framework and experimental setup were designed together. The author implemented the retrieval system and carried out the experiments. Writing of the paper was a collaborative effort.

1. Introduction

The importance of information retrieval (IR) keeps growing as the amount of digital information keeps expanding at an ever-increasing rate. Stored documents, photographs and contents of books and billions of web pages are useful only if they can be found when needed. Web search engines are the most common common way to find information. They are attracting more than 170 billion queries each month [Bonfils, 2013].

While the current search engines are undoubtedly useful, they also have their limitations. Search engines typically expect the query to be represented as a few search terms. It may be difficult to express complex information needs this way [Spink et al., 1998]. Sometimes constructing a search query may be too laborious, or one may not even be aware that valuable information exists and could be found by searching.

An emerging research trend aims to address these problems. A new generation of intelligent search engines employ machine learning to better understand user's goals. The search engines record how users interact with them and learn from the collected past observations to return more relevant results [Joachims and Radlinski, 2007]. The overall goal is to automatize part of the search process so that less actual effort is required from the user.

An intelligent search engine has at least the three main parts shown in Figure 1.1. To be able to reason about and adjust to user's needs the system must observe and collect data about how the user interacts with the search engine. This *feedback* data can be collected by directly asking the user to rate some documents, by instrumenting applications to record indirect behavior patterns or by external sensors, such as eye trackers. Mobile devices can provide further useful data about the location and the surrounding environment. Indirect measurements are valuable because they can be collected passively without distracting the user.

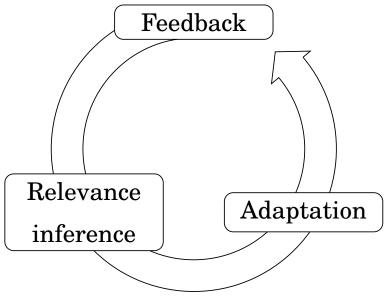

Figure 1.1. Main components of an intelligent search engine: User's actions are recorded and interpreted as feedback for the system. Relevance inference methods estimate user's preferences. Interface and search engine behavior are adapted to better match the estimates. Adaptation leads to new options for the user who provides more feedback by interacting with the updated options.

The feedback contains valuable information about user's preferences and goals. *Relevance inference* methods extract this information from the collected feedback data. They are statistical and machine learning techniques that are used to build models of user state, intent or context [Kobsa, 2007]. The learned models provide estimates of which items or actions the user prefers. These estimates can represent or simulate the user so that the system can make some choices automatically on the behalf of the user.

The feedback data is often very noisy; most interactions are not directly related to relevance but reflect user's behavior when she is exploring the options and thinking about what to do next. Statistical methods are required to uncover useful signals from the recorded data.

The most important advantage of the relevance inference is that the system can *adapt* its interface and behavior to user's preferences. A search engine can, for example, show personalized search results or correctly disambiguate search terms if it is able to estimate user's interests. The adaptation benefits the user by lower the number of interaction steps required to complete a task or by presenting information in a format that is easy to understand because it matches user's mental model.

After adaption, new options, which are estimated to be relevant, are presented to the user who makes a selection among them or otherwise interacts with them. The new interactions are again captured and interpreted as feedback for further tuning of the relevance estimates. This turns the whole process into a loop which, in the ideal case, continuously improves the quality of the displayed results.

There are many problems in making search engines more intelligent, but the challenges are worth confronting because the result will be faster and more effortless access to relevant information. New search methods and principles even enable completely new ways of accessing information.

1.1 Objectives and scope

This thesis studies intelligent information retrieval methods and applications. Specifically, the thesis introduces novel machine learning methods for relevance inference and studies two cases of adaptation in IR applications: *proactive information retrieval* applications automatically suggest relevant content, the second approach is providing *browsing cues* by making relevant items more salient and, thus, easier to recognize.

The machine learning contributions in this thesis focus on inferring relevance in situations where little feedback is available for the current search session but feedback has been been given on similar search tasks before. Publications I, II and VI study how feedback from different but related search sessions can be combined in order to reduce the problems caused by sparsity of feedback in any single search session. The feedback from different search sessions can not be directly merged because the search goals are different if the searcher was, for example, looking for information about cats in one session and about dogs in another. It is shown that some evidence can still be transferred reducing the amount of feedback required for accurate relevance prediction. The main objective of Publications I and II is to explore novel methods for combining feedback collected from different search sessions. In these publications, feedback is constituted by reading patterns recorded by an eye tracker. Publication VI studies how to combine feedback given for multiple types of data, for example texts and images, in a single search task.

Novel proactive information retrieval applications are presented in Publications I, II and III. These Publications describe automatic retrieval of documents that are relevant to user's current needs. The applications presented in these publications infer relevance from implicit eye movement patterns and do not therefore require any more cognitive effort than what is normally spent on observing the stimulus. Publications I and II describe a search engine that infers query terms based on reading patterns. The application in Publication III turns the physical environment into an information browser by merging search results into the real world through an augmented reality display. Implicit feedback for both

Table 1.1. Publications and their contribution areas.

	Publications					
	I	II	III	IV	V	VI
Relevance inference methods	X	X				X
Proactive information retrieval	X	X	X			
Browsing cues				X	X	
Novel search applications:						
Query prediction from reading patterns	X	X				
Real word information access			X			
Personal history browser				X	X	X

the physical environment and the augmented content is captured by a wearable eye tracker.

The second type of adaptation technique studied in the thesis is providing browsing cues to help to understand the context of search results. If the data can be arranged in an meaningful order, for example along a timeline, data items can act as orienteering points for deciding which direction to explore next. Emphasizing relevant items as potential browsing cues is expected to aid the user in the exploration. This works best if the data is at least somewhat familiar to the user, such as a personal history. Publications IV and V study how much browsing cues improve retrieval performance in the context of a personal history browser application. The interfaces with browsing cues are compared against more traditional interfaces that do not emphasize relevant items. The proposed methods select potential browsing cues based on their inferred relevance and increase their salience to make them easier to recognize.

Table 1.1 summarizes the relationships between the Publications and the contribution areas. Prototype versions of all discussed applications have been implemented and validated in user studies.

This thesis focuses on small-scale retrieval applications. The field of information retrieval research is extensive, and many topics, such as scaling up to web scale and effective indexing techniques, are outside the scope of the thesis. There exist many other information sources, such as user profiles and behavior of other users in similar situations, physiological and other measurements, that may be useful in relevance estimation, but those are not discussed in this thesis. Neither are privacy implications of recording and storing user feedback, although it is an important theme. The thesis focuses on IR applications, but the idea of learning and adapt-ing to user's preferences is more general. A similar approach has been proposed, for example, in designing adaptive user interfaces [Kandel et al., 2011].

The thesis is structured as follows. Chapter 2 provides a summary of literature on information retrieval research related to the themes in the rest of the thesis. Chapter 3 discusses machine learning methods for sharing relevance predictions among retrieval tasks, with applications to constructing queries proactively from reading patterns. Chapter 4 discusses merging of retrieval results with the real world through augmented reality. Browsing cues and the personal history browser are presented in Chapter 5. Chapter 6 reports work on learning relevance by transferring feedback from other data domains.

2. Information retrieval

The field of information retrieval (IR) studies computer systems for finding documents that satisfy user's information needs. The retrieved items may be either textual documents, images, videos or other types of digital content. The scale of the information retrieval extends from finding a file on a personal computer to web search engines that search among billions of documents stored on tens of millions of networked computers.

This Chapter reviews previous research on the research ares related to the topics of this thesis. Section 2.1 presents the standard keyword-based search and its limitations. The following Sections introduce techniques that address some limitations of the keyword search. Section 2.2 explains how relevance feedback reduces the reliance on search terms. One source of relevance feedback is eye movements as elaborated in Section 2.3. The thesis studies two applications for relevance inference: browsing cues and proactive information retrieval. The role of browsing cues in retrieval is discussed in Section 2.4. Section 2.5 introduces proactive information retrieval systems that automatically provide necessary information when needed.

This Chapter does not aim to be a comprehensive overview of IR research; many important areas, such as retrieval of linked, structured and multimedia data, are not discussed. A more exhaustive treatment is provided by Manning et al. [2008].

2.1 Keyword search

To use a digital IR system, one must specify the information need in a form that is understandable to a computer. The most common search interface is keyword-based search. The user enters search terms, and the search engine retrieves from its index documents containing the given terms or their synonyms. The search engine typically ranks the documents in the order of estimated relevance before displaying the results. A major application area for keyword search is web search.

A common implementation technique is the so-called *vector space model* [Salton, 1971]. Both the documents and the query are presented as numerical vectors in a concept space. The relevance ranking is formed by ordering the documents according to their similarity to the query vector in the concept space.

Despite its popularity, the keyword search is not always the most accurate or the most pleasant way of accessing information. It is well known that formulating good textual queries is difficult. Novice users find keyword search confusing [Hargittai, 2004], and would often prefer natural language queries [Bilal, 2000]. Even for experienced users constructing a query can be challenging, for example, when the information need is vague [Spink et al., 1998], or because the searcher does not know what kind of information is available [Salton and Buckley, 1990], or if there is a mismatch between the vocabularies used by the user and the system [Belkin, 2000].

2.2 Relevance feedback

The standard search process can be extended to include a feedback phase. Giving feedback on the documents may be cognitively less demanding than formulating good search terms. Therefore, an ability to provide feedback has been proposed as a user friendly way to guide the search.

Relevance feedback [Salton, 1971] is an iterative search process that consists of interaction between the searcher and the search engine. As a first step, the searcher enters an initial query, which may be vague. As a response, the search engine returns a list of documents. The searcher can view the result titles and snippets and open some of the documents for closer reading. After inspecting the initial set of results, the searcher marks some documents as relevant and some as not relevant and instructs the search engine to retrieve more documents. The next set of result documents is retrieved. The search engine chooses the new documents so that they are similar to both the initial query and the documents labeled as relevant. Because more information is available to the search engine, the new documents should better match searcher's intent. The process continues until the searcher finds a satisfactory document or gives up.

In the vector space model, the relevance feedback is implemented by moving the original query vector towards the feedback documents. The updated query is a weighted vector sum of the original query vector, the relevant document vectors and negated non-relevant document vectors. The documents most similar to the updated query vector are retrieved and returned as the result.

The relevance feedback may be *explicit* or *implicit* in nature. Explicit feedback can be collected with a search interface that allows rating the search results as relevant or non-relevant, or on a more refined scale. Implicit feedback consists of user interaction observations that can be collected passively while the user is working with the documents. For example, the duration one spends reading a page can be used as an implicit feedback measure. Explicit feedback is often more accurate but also more laborious to assign, while implicit feedback is more noisy but also essentially free to collect during normal operation with no additional effort required by the user.

Explicit relevance feedback is known to improve the retrieval accuracy [Salton and Buckley, 1990]. However, the relevance feedback features are not commonly used even when available [Ruthven et al., 2001], and the major web search engines have been reluctant to adopt them. A possible explanation is that assessing the relevance of a potentially complex document is too demanding, and the users are unwilling to do it, if they do not see immediate benefits. Including the options to give feedback may also make the user interfaces too complicated [Aalbersberg, 1992].

Implicit feedback is inferred from various interaction patterns, for example, if a hyperlink is followed [Joachims and Radlinski, 2007], the duration of time a document was displayed [Kelly and Belkin, 2004] or if a document is saved [Morita and Shinoda, 1994]. Kelly and Teevan [2003] reviewed previous research on implicit feedback and concluded that, while the implicit feedback can be useful, it is not necessarily so. Careful modeling of the user's goals is required to correctly interpret implicit feedback. Fox et al. [2005] reported that a properly selected combination of implicit measures correlates with explicit satisfaction ratings. Teevan et al. [2010] used implicit feedback to estimate how much search engines would benefit from personalization instead of presenting the same results for everybody.

According to White et al. [2005] implicit relevance feedback is generally preferred in complex search tasks and explicit feedback in simple tasks.

In complex tasks, finding fully relevant documents is difficult and, therefore, the searchers need to spend considerable effort in judging the usefulness of documents as relevance feedback. The same study also found out that inexperienced users especially like implicit feedback, because they are less familiar with the explicit feedback process.

2.3 Eye movements as implicit relevance feedback

Eye gaze location is a good indicator of moment-to-moment focus of attention. Consequently, eye tracking is an interesting source of implicit feedback.

2.3.1 Physiology of the eye

The area of sharp vision is only 1-2 degrees, and therefore the eyes must move around to gather details from a scene. The eye gaze location is often closely related to the focus of attention (in rare cases the attention may be directed away from the point-of-gaze). The eyes move by irregular, rapid jumps called *saccades*. Between the saccades the eyes stay almost stationary on one point. These periods of immobility are called *fixations*. Fixation durations vary between 50 and 500 ms depending on the task and the complexity of the scene. The saccades are much shorter, typically only 20–50 ms. A sequence of fixations and saccades forms a *trajectory*. An example trajectory is plotted in Figure 2.1.

The neural processes that decide where to look next, that is, the processes responsible for planning and execution of saccades, are partly controlled by intrinsic scene saliency, derived from contrasts in color, shape or motion, and partly by mental model of potentially informative locations built during the scene viewing so far [Kowler, 2011]. The controlling process depends heavily on the task.

During reading, saccades are shorter than in image viewing. Fixations mostly advance from left to right, but about 10%–15% of saccades are regressions back to previous words. Some words are skipped completely and others are fixated several times. The frequency of skipping and regressions depend on the difficulty of the text and the mental state of the reader. Rayner [2009] provides a comprehensive review of characteristics of eye movements during reading.

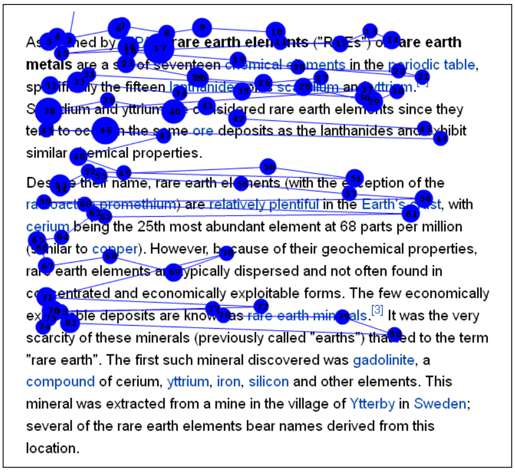

Figure 2.1. Eye movement trajectory during reading. The fixations are marked by circles and the saccades by the connecting lines. The radius of the circle is proportional to the fixation duration.

2.3.2 Eye tracking devices

Many kinds of devices for recording eye movements have been developed [Duchowski, 2007]. The most common ones are video-based methods, which detect the position and direction of eyes from a video feed captured by a camera. Other options include special contact lenses, whose movements cause measurable magnetic forces, and electrooculography (EOG) sensors attached to the skin around the eye. The EOG sensors detect weak electric signals caused by the movement of the eye. The location of gaze on a computer monitor or other planar surface can be measured quite easily using a remote video-based eye tracker without having to wear contact lenses or head gear. Modern eye trackers tolerate natural head movements during measurement so that there is no need to stabilize test subject's head. Wearable eye trackers extend the eye tracking capabilities from computer-bound scenarios to natural situations.

In the beginning of a recording session the eye tracker must be calibrated so that the pupil direction can be mapped to corresponding screen coordinates. The calibration procedure involves viewing a sequence of calibration points successively. The measurement device software is able to compute the calibration parameters based on the measured eye directions and the known coordinates of the calibration points.

While commercial eye trackers are still too expensive for most consumer applications at the moment, some researchers have experimented with building eye trackers from low-cost components [Agustin et al., 2010, Mantiuk et al., 2012]. The prices have been decreasing and the quality has increased during recent years. This trend is promising for eye trackers to become more common also outside research laboratories.

2.3.3 Applications of implicit eye movements

The most direct gaze-interaction technique simulates the mouse; the cursor location is controlled by eyes and activation or selection is done by staring at a object for a certain time (usually around 500–1000 ms) [Pfeiffer et al., 2008, Park et al., 2008, Nilsson et al., 2009] or by gaze gestures, such as blinking [Ishiguro et al., 2010, Baldauf et al., 2010, Lee et al., 2011]. Such gaze-control has been used in text entry through on-screen keyboards (e.g. [Hansen et al., 2001, Fejtová et al., 2004]). However, these kinds of interfaces are cumbersome because the eyes are normally used for observation, not for control.

In this thesis, the main interest is on eye movements as implicit feedback. Instead of controlling gaze location consciously, the viewer looks at the scene in a natural way. The burden of separating purely observational eye movement patterns from patterns that relate to real interest is left to a machine learning algorithm [Bednarik et al., 2012].

Inferring user goals from eye movements seems plausible, because eye movement patterns change depending on cognitive state. For example, Velichkovsky et al. [2002] showed that fixation durations generally become longer when the level of mental processing increases. Similarly, pupils are known to dilate with increased mental workload [Pomplun and Sunkara, 2003]. These and other more complicated patterns can be detected using machine learning methods. These kinds of low-level patterns can act as building blocks in predicting higher-level intentions. For example, Simola et al. [2008] predicted which of three types of search tasks the user was performing based on the eye movement trajectory.

The task of making inferences based on an implicit eye movement trajectory can be conceptualized as relevance prediction; an algorithm is trained to use eye movements in predicting which sections of text, objects in an image, or other targets-of-interest in a scene are relevant. Salojärvi et al. [2003] studied relationships between various eye movement features and text relevance. Later, Salojärvi et al. [2005] trained an eye movement based relevance prediction algorithm for a text retrieval task. Puolamäki et al. [2005] extended the predictor to include user preference ratings in addition to eye tracking. Buscher et al. [2008] introduced an algorithm for inferring sub-document relevance from eye movements and used the identified document sections for query expansion in IR. Recently, many researchers have shown interest in using eye movements as implicit feedback in image retrieval tasks [Oyekoya and Stentiford, 2006, Klami et al., 2008, Auer et al., 2010, Hardoon and Pasupa, 2010]. Eye movements have also been used to detect relevant scenes in videos [Kandemir et al., 2010].

Attentive documents react to how they are being viewed. An early example was an interactive story teller that modified the story according to which objects in images were viewed [Starker and Bolt, 1990]. Eye tracking can be used to detect which parts of a text document are read thoroughly. When the document is reopened, the read segments can be highlighted to help recognizing the document or to emphasize the relevant locations [Ohno, 2004, Buscher et al., 2012]. Alternatively, Xu et al. [2009] proposed using the read segments in generating personalized document summaries. Hyrskykari et al. [2003] introduced an interactive gaze-based assistant that infers from eye movement patterns when a reader has difficulties in understanding foreign text, and offers a translation. Interactive books react to reading by playing appropriate sounds or animations when the eyes reach a pre-determined location in the text [Biedert et al., 2010].

2.4 Searching by navigation and browsing cues

Several studies have shown that people prefer navigation over a focused keyword search when searching in personal emails, files and on the Web [Barreau and Nardi, 1995, Teevan et al., 2004, Boardman and Sasse, 2004]. According to Bergman et al. [2008], keyword search is often a last resort in file search tasks. It is used only when the user does not remember the location of the file. This observation holds despite the existence of advanced desktop search engines.

Searching by navigation is an iterative process where one takes small steps towards the target. Examples of navigation behavior include locating information by following promising-looking web links, applying a new filter in a search interface or updating search terms after seeing results of a query.

Reasons for the navigation preference are numerous. Proceeding towards a target in small steps may be cognitively less demanding than a keyword search, because coming up with search terms is difficult [Bergman et al., 2012]. Navigation may be less expensive also if one remembers the steps to reach the target from a previous search session [Bergman et al., 2008]. Another benefit of narrowing down towards a target in a series of small actions is that the process helps to understand the context of the answer [Teevan et al., 2004].

Contrary to classical assumptions in information retrieval research, the information need often changes during a search session [Bates, 1989]. The need to be able to alter the search is another reason to provide navigational features in search interfaces. The query may be evolving because the searcher may focus or broaden it after seeing some initial results, because new associations may pop into mind, or because the searcher may be able to formulate a query in a way that matches the document representation only after familiarizing herself with the contents of the database.

The dynamic nature of an information need can be taken into account by weighting recent relevance feedback more than older feedback [Campbell, 2000, Hopfgartner et al., 2008]. This means that, after the information need changes, feedback received before the change is given gradually less weight and will eventually be forgotten. Alternatively, the search engine may estimate the level of change in the information need and make comparable changes in the interface [White et al., 2006]. If an action is estimated to change the information need only moderately, the engine may choose to just re-rank the top documents, whereas if the change is substantial, the engine may execute a completely new search.

YouPivot [Hailpern et al., 2011] is an example of a retrieval application that promotes navigation. It records simultaneous activity, such as documents being read and songs being listened to, and enables browsing of events that are aligned in time. One can find a document even when one only remembers which song was playing while one saw the document, by asking the system to list all documents that were open when the particular song was last heard. GaZIR [Kozma et al., 2009] is an example of navigational interfaces on image domain. One can zoom in to the image collection to see images that are similar with each other or backtrack to a previous state in order to start following a different path. The relevance is estimated based on what kinds of images have drawn attention as measured by an eye tracker.

Searching by navigation is more effective if one can easily understand the current state of the search. One way to promote fluent understanding is presenting *browsing cues* to help a searcher to orient herself. The cues are something that are readily recognized and indicate what kind of items are being displayed nearby. An example is a search interface by Ringel et al. [2003]. It shows search results ordered on a timeline aligned with such cues as selected public news headlines, personal calendar events and photographs. The cues aid in recognizing which part of the history is being shown more than simple dates would, because people are more likely to remember relationships between events than exact dates. Ringel et al. found out that searching with the cues enabled is significantly faster than without them. Doherty and Smeaton [2010] suggested annotating a personal data search result display with content from image and video sharing web sites.

All items do not make equally good browsing cues. Items related to memorable moments are more likely to be recognized long after the event and should therefore be better cues. Horvitz et al. [2004] proposed a probabilistic inference model for predicting which calendar events are remembered well. In a study with data collected from real users, their model identified atypical duration or location and non-recurrence of an event as good indicators of a memorable event. Later, Hwang and Cho [2009] adapted a similar inference process to run on a mobile phone with limited computing resources.

If the user is already familiar with the data, it is also possible to utilize the data items themselves, instead of external content, as browsing cues. The items that are estimated to be good navigation cues can be emphasized relative to others. Lee et al. [2008] showed images sized proportionally to their visual uniqueness on the assumption that unique images are better cues than commonly occurring images. PhotoMemory [Elsweiler et al., 2007] enlarged images matching given visual, temporal or other metadata filter values.

2.5 Proactive information retrieval

The most radical departure from the standard keyword search paradigm is *proactive information retrieval*. In general, a proactive system aims to predict user's goals and offers relevant suggestions. Examples of proactive systems include text input systems that predict likely continuations using language modeling [Ward et al., 2000] and adaptive user interfaces that suggest next actions based on the current state [Kandel et al., 2011, Wolfram, 2012].

A proactive information retrieval system infers what kind of information the user wants to see next, performs the search automatically and has the information ready when the user needs it. The prediction is based on (often implicit) feedback and inference models learned from past observations.

The main component of a proactive system is an inference engine that predicts what are the possible actions the user might want to take next given the application state and steps taken so far. Typically, a proactive system generates a list of candidate actions, ranks them based on how likely they are at the current state, and displays the top suggestions. The inference rules can either be constructed by application experts or learned by collecting a large pool of data on which actions users tend to take in specific situations.

Proactive search has some important advantages over keyword-based search. Because the retrieval process is largely autonomous, no effort is spent in constructing search terms. A proactive IR system is able to perform search queries in the background while the user is engaged in other activities. When the user interrupts the work to ask for more information, the system already has relevant recommendations ready. Finally, automatically generated recommendations can expose relevant information even when the user does not know it exists [Laqua et al., 2011].

The proactive recommendations should be displayed in an accessible yet non-distracting manner. If the user interface contains visible recommendations that are constantly changing in response to user's actions, they can distract the user from the main task. Some ways to maintain the usability include increasing saliency of the recommendations only when the prediction is very confident or showing the recommendations only when the user explicitly asks for them [Billsus et al., 2005]. The system can also try to predict a suitable time for showing notifications. For example, interrupting the user at sub-task boundaries has less negative effects on task performance [Bailey and Iqbal, 2008]. However, if the urgency of the notification is estimated to be high enough, the system might still choose to show it immediately [Grandhi and Jones, 2010].

One goal of proactive IR research is an automated personal assistant that knows user's preferences and current context and is able to give relevant answers to complicated queries. This line of research was explored by DARPA CALO project [SRI International] and is now being commercialized by Apple's Siri[1], Google Now[2] and other similar mobile applications. Other examples of proactive IR are writer assistants that provide recommendations based on the text being written [Rhodes and Starner, 1996, Rhodes, 2000, Ramos et al., 2008]. They constantly monitor what part of the text the user is currently editing, submit related search terms to a search engine, and display results on the text editor window. Alternatively, proactive queries can be constructed based on read emails [Dumais et al., 2004, Laqua et al., 2011].

Contextual search [Lawrence, 2000, Jones and Brown, 2004, Kraft et al., 2006, Adomavicius and Tuzhilin, 2011] is a closely related research field. Unlike purely proactive retrieval, contextual search is still initiated by the user entering search terms. Observed context is used to refine or to disambiguate the query. The context can include location, the text content of recently displayed web pages and other documents [Gyllstrom and Soules, 2008], previous queries [White et al., 2010], file access patterns [Soules and Ganger, 2005], document creation time and location [Fuller et al., 2008], or other similar data sources.

A major challenge in proactive IR is accurate inference based on noisy and sparse (implicit) feedback and contextual features. Robust statistical models are needed for the inference. This thesis proposes two inference methods that improve the prediction by combining feedback from related tasks. Another challenge, which is outside the scope of the thesis, is modeling user's activity. The relevance of recommendations can presumably be improved, if user's goals are modeled properly.

[1]http://www.apple.com/ios/siri/

[2]http://www.google.com/landing/now/

3. Learning implicit queries from gaze patterns

This Chapter introduces machine learning methods for inferring queries from eye movement feedback. The main research questions are if it is possible to infer mental queries form eye movements and how to combine feedback from different search sessions. The methods are studies in a context of a proactive IR application. The methods and the application were originally studied in Publications I and II.

3.1 Implicit queries

A proactive information retrieval system typically includes a component that constructs query terms based on user's actions and submits them to a retrieval engine. This component tries to anticipate what kind of queries the user is likely to submit after executing certain actions. If the query is derived from implicit feedback, it is called an *implicit query*. An implicit query is constructed automatically, without explicit commands, while the user is working, and the retrieval results matching to the query are displayed as recommendations related to the user's current work.

This Chapter describes how implicit queries can be estimated from eye tracking measurements while a person is reading text documents. Two variants of a machine learning method for estimating how important a reader found a word during reading are introduced. The words and their estimated importance weights can be considered as a weighted query, which can be submitted to a search engine to retrieve documents containing many terms whose estimated weight is high and few terms whose estimated weight is low.

Others have previously constructed implicit queries based on the text a user is editing in a word processing application [Rhodes and Starner, 1996, Czerwinski et al., 1999, Dumais et al., 2004]. A natural extension is to use eye movements to detect which *parts* of the text are relevant [Maglio and Campbell, 2003, Salojärvi et al., 2005, Puolamäki et al., 2005, Buscher et al., 2008, 2012].

Several modeling approaches for detecting relevant parts of text based on eye movements have been tested. The simplest weight the words by the viewing duration [Miller and Agne, 2005]. The duration, however, offers only a partial and imperfect view of the relevance. The way eyes move between neighboring words is also informative. Others have built more complicated rule-base prediction models by hand [Buscher et al., 2008, 2012, Umemoto et al., 2012]. We extend these methods by providing a machine learning method that infers the relationship between eye movements and relevance from data.

Salojärvi et al. [2005] and Puolamäki et al. [2005] were the first to use machine learning to predict relevance from eye movements. They trained discriminative Hidden Markov Models to predict from an eye movement trajectory which lines in a list of displayed news titles are relevant. Loboda et al. [2011] used a linear mixed-effects model to analyze how different eye movement features contribute to word relevance.

3.2 Learning queries from eye movements

This Section describes how implicit queries can be learned from eye movements. Although the main goal is to learn from implicit feedback, some explicit feedback must be available at some point before the method can learn the link between the eye movements and the relevance. To this end, it is assumed that the user has completed other *search sessions* before. During those sessions, explicit document relevance labels as well as eye movements have been collected. For example, in the user experiments we conducted, the test subjects were looking for information about education in one session and information about the Olympics in another session. The assignment of documents and eye movements into search sessions is assumed to be given, that is the session boundaries are not estimated from the data.

The usage scenario is that the user first does some searches, reads documents returned by the search engine, and marks them as relevant or not relevant. The eye movement data is recorded also. After enough training data has been collected, the method can be applied in a proactive mode on new search sessions. On a new session, the user just reads documents,

Table 3.1. The training data consists of binary relevance labels (+ = relevant, - = not relevant) for a subset of documents on a few search sessions and eye movements on those sessions. The method's goal is to predict document relevances on the test session given only eye movements on that session and all data on the training session.

	Eye movements available?	Document relevance labels						
		d_1	d_2	d_3	d_4	d_5	d_6	$\dots$
Training session 1	yes	+	+	-				$\dots$
Training session 2	yes			+	+	-		$\dots$
Training session 3	yes		-		+		+	$\dots$
$\vdots$								
Test session	yes				?			

without labeling them, to find information about some topic, which may be different from all topics in the training sessions. After some eye movements have been collected on the new session, the model is able to estimate an implicit query, which is used to find new documents to read. Table 3.1 shows what data is collected in the various sessions.

The machine learning method uses the all collected relevance labels and eye movements to learn a connection between eye movements and importance of a word. The relevance labels can not be readily pooled, because the information needs are different. For example, a document that discusses education is unlikely to be relevant when looking for information about the Olympics. However, it is possible to integrate data from different sessions by assuming that certain characteristics of eye movements do not change between sessions.

To investigate the feasibility of using eye movements as a way of sharing feedback across sessions, we make an assumption that there is a connection between the importance of a word and the way the word is viewed. We further assume that this link function is universal, that is, not specific to individual search sessions. In other words, the assumption is that words that are relevant for an information need are always viewed in a particular way, regardless of what the information need is, and that irrelevant words are viewed in a different manner. It is known that one typically reads relevant sentences or paragraphs more carefully than uninteresting parts [Duggan and Payne, 2011]. It seems plausible to assume that these kind of differences in reading style generalize across search sessions.

If the assumption holds, the link function form eye movement to word importance can be learned on training sessions, where ground truth about document relevance is available. The learned link function can then be applied on a new search session, where only eye movements have been observed, to estimate word importance values. These values are estimates of how important the reader implicitly thought the viewed words were during reading. They are fed to a search engine that ranks unseen documents according to how similar their content is to the estimated implicit query. The first few documents in this ordering are displayed to the user as recommendations.

3.3 Related work on machine learning

This Section gives an overview of research on related areas of machine learning.

3.3.1 Multitask learning

Multitask learning approach to machine learning aims to improve generalization performance by sharing model parameters over multiple related tasks [Thrun, 1996, Caruana, 1997, Giraud-Carrier et al., 2004]. When a model learns several tasks in parallel, the tasks act as inductive biases for each other. Our assumption that the reading style on relevant words is similar on different search sessions naturally leads to such representation: the parameters describing the reading style are shared between the sessions.

Learning relevance function over multiple search sessions has previously been studied in the context of feedback extracted from search engine log data [Joachims, 2002, Agichtein et al., 2006]. These works learned one static function that was then applied to all sessions. We instead learn session specific classifiers that share some parameters.

3.3.2 Support vector machines

In Publication I, we employ support vector machines (SVMs) [Cortes and Vapnik, 1995] as part of our model. SVMs are general classifier functions that seek for a classification boundary as a hyperplane that maximizes the margin among the classes on the training data.

The optimal hyperplane w of a linear SVM is the solution of the following minimization problem:

$$\min_{\boldsymbol{w},b,\xi_i} \left\{ \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{N} \xi_i \right\} \tag{3.1}$$

$$\text{subject to} \quad y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i - b) \geq 1 - \xi_i, \xi_i \geq 0, \text{for all } 1 \leq i \leq N,$$

where $\boldsymbol{x}_i$ are the training samples, y_i are their class labels ($+1$ or -1), ξ_i are slack variables, and C is a trade-off parameter between a large margin and a small error.

3.3.3 Bayesian inference

In Publication II, we formulate an alternative solution to our problem using Bayesian generative modeling. In Bayesian inference [Gelman et al., 2004], uncertainty in parameters $\boldsymbol{\theta}$ is modeled by a joint probability distribution. The inference proceeds by computing the influence of observed data $\boldsymbol{X}$ to the distribution of parameters $\boldsymbol{\theta}$ using Bayes' theorem:

$$p(\boldsymbol{\theta}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{X})}, \tag{3.2}$$

where $p(\boldsymbol{X}|\boldsymbol{\theta})$ is the *likelihood* or the probability of the observation given a parameter value; $p(\theta)$ is the *prior probability* or the probability of parameters before observing any data; and the marginal likelihood $p(\boldsymbol{X}) = \int p(\boldsymbol{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$. Except in very simply cases, the Bayesian models are usually hierarchical in the sense that the prior distribution is further decomposed into conditional distributions: $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)p(\boldsymbol{\theta}_2)$. The conditioning parameters $\boldsymbol{\theta}_2$ are called *hyperparameters* and their distribution $p(\boldsymbol{\theta}_2)$ a *hyperprior*.

A major computational issue with the Bayesian inference is that the computation of the marginal likelihood $p(\boldsymbol{X}) = \int p(\boldsymbol{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$, or other integrals that arise, is often intractable. Several approximation methods have been introduced.

3.4 Models and results

This Section describes two models for the task of constructing implicit queries from eye movements and the results of user studies conducted to test them in an information retrieval setting.

We build a set of hierarchical text classifiers, one classifier for each search session. The hierarchical structure is shown in Figure 3.1. The classifiers are trained to assign binary relevance labels (relevant or not

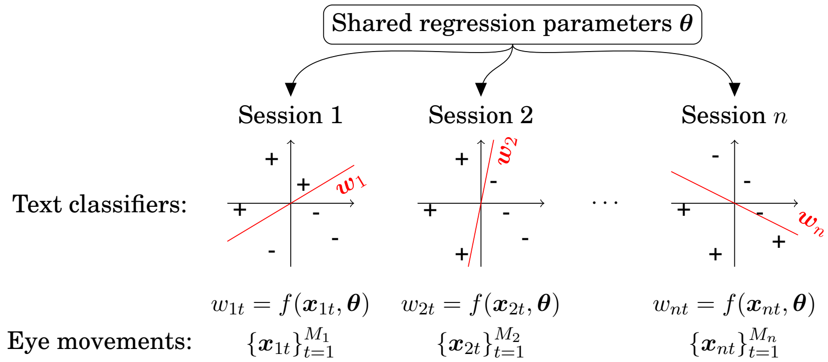

Figure 3.1. Session-specific text classifier separate documents into relevant (+) and not relevant (-). The classifier parameter vectors $\boldsymbol{w}$ depend on eye movements $\boldsymbol{x}$ during the session and the shared regression parameters $\boldsymbol{\theta}$.

relevant) to documents. The classifier for session k is parametrized by a latent word vector $\boldsymbol{w}_k$, where each component corresponds to one word in the vocabulary. According to our assumption these latent word weights w_{kt} depend on the recorded eye movements $\boldsymbol{x}_{kt}$ on the word t during session k. We assume that the dependency takes a form of a regression function $w_{kt} \sim f(\boldsymbol{x}_{kt}, \boldsymbol{\theta})$. The parameters $\boldsymbol{\theta}$ are shared across all sessions. This sharing makes it possible to generalize from the training sessions to the test sessions.

The eye movements on a particular word are summarized as a feature vector. The features consist of eye movements on the word, such as number of fixations and total fixation duration on the word, eye movements between the neighboring words, such as length of a saccade when entering or leaving the word, and regressions back to earlier words, such as number of regressions starting from a word. In addition to the eye movement features, we use a few textual features, such as inverse document frequency and relative position of the word on a line. The textual features are independent of the semantic meaning of the word and therefore can be assumed to be useful in generalizing among the sessions.

The specific training procedure depends on specific model structures selected for the roles of the text classifier and the word importance regression. In the Subsections below, we discuss two possible choices.

3.4.1 SVM model

Publication I uses support vector machines (SVMs) as session specific text classifiers and a linear and non-linear regression functions for word importance regression.

The learning consists of two steps: first, an SVM is learned for each training session separately to discriminate between relevant and non-relevant documents, then a regression function is trained to estimate the relationship between the eye movement features and the learned SVM parameters. As discussed above, the regression parameters will be shared across the sessions.

More formally, the training instances for the regression function are pairs of eye movement feature vector $\boldsymbol{x}_{kt}$ for term t on training session k and learned SVM parameter value w_{kt} on the same term and training session. A regression function f, parametrized by $\boldsymbol{\theta}$, is learned by minimizing the squared error $(f(\boldsymbol{x}_{kt};\boldsymbol{\theta}) - w_{kt})^2$ over all terms t and training sessions k. If different instances of a word t are viewed in a session k, the corresponding feature vector $\boldsymbol{x}_{kt}$ is an average of feature values observer on the different instances.

In principle, any regression algorithm can be applied. We tested both the standard linear least squares regression and a non-linear sparse dual partial least squares method [Dhanjal et al., 2006].

Applying the learned model to a new retrieval session has three stages: collecting eye movement features $\boldsymbol{x}$ when a person is reading documents with an information need in mind, estimating word relevance scores using the regression function $f(\cdot;\boldsymbol{\theta})$, with the $\boldsymbol{\theta}$ value that was learned before, and classifying documents with an SVM, whose parameter vector $\boldsymbol{w}$ consists of the estimated word relevance scores: $w_t = f(\boldsymbol{x}_t;\boldsymbol{\theta})$. Both of the computational steps are fast when using a linear regression and a linear SVM, because only linear-time operations are involved.

Experiments

We evaluated the performance of the implicit query prediction SVM model in a series of controlled experiments. Eye tracking data was collected from test subjects who read short text snippets trying to identify if they were about a given topic. Each test subject completed several sessions.

The testing was done in a leave-one-session-out fashion: document relevances on one session were hidden, and an implicit query was estimated using the document relevance and eye movement data from the other sessions. Documents on the test session were ranked according to the implicit query, and the resulting ordering was compared to a known ground truth.

The model learned using the eye movement and textual features performed better than a similar model that used only textual features. This implies that there indeed exists a dependency between eye movements and interest, confirming our assumption. Furthermore, it is possible to utilize this dependency to do proactive retrieval by implicit eye movement feedback.

An interesting question is which eye movement features help in discriminating between relevant and non-relevant words. We analyzed the regression coefficients of the linear model and found three coefficients that differ significantly from zero: saccade length before first fixation to the word, indicator variable for a regression being initiated from the next word and duration of the first fixation to a word relative to the total fixation duration. In other words, both how long a word is being viewed (relative to other words) and the style of saccadic movement between the neighboring words are indicative of the relevance of the word. The features used in our study have been initially proposed in various psychological studies, not necessarily related to information retrieval. It is likely that it would be possible to construct features that are even better suited for the relevance prediction task.

3.4.2 Probabilistic model

Publication II proposes an alternative model to the same problem which was studied in Publication I. The new model avoids a particular difficulty in the training of the previous model.

The training of the model discussed in Subsection 3.4.1 composed of two steps: learning the text classifiers and the regression function. Having two separate steps seems suboptimal. In Publication II we show that it is possible to construct a single probabilistic model that incorporates both steps. Our hypothesis is that combining the two steps into a combined optimization procedure should lead to a better overall performance.

The experimental setting is similar as in the previous Subsection; eye movements and relevance feedback are recorded while subjects are reading documents, and the model is trained to estimate an implicit query on a session, where only eye movements are available.

We construct a generative hierarchical model for predicting document relevance. The plate diagram of the model is shown in Figure 3.2. A logistic regression function is trained to predict relevance r of a document:

$$p(r = 1|\boldsymbol{d}, \boldsymbol{w}) = \frac{1}{1 + e^{-\boldsymbol{w}^T \boldsymbol{d}}}, \tag{3.3}$$

where $\boldsymbol{d}$ is document vector, and $\boldsymbol{w}$ is a latent query parameter vector,

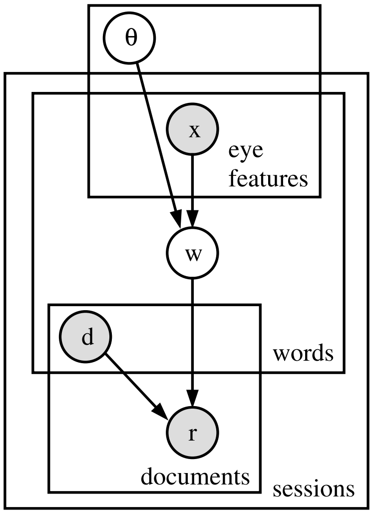

Figure 3.2. A plate diagram showing the dependencies between the variables in the probabilistic model for predicting document relevance r given the document term vector $\boldsymbol{d}$ and a latent word weight vector $\boldsymbol{w}$. The latent parameters $\boldsymbol{w}$ depend on observed eye movement features $\boldsymbol{x}$ on the same session and on regression parameters $\boldsymbol{\theta}$ that are shared across all sessions. The shaded nodes are observed.

with one dimension per each word in the vocabulary. The query vector has a hyperprior that depends on eye movements on the read documents. Each component is assumed to be normally distributed around a mean that is a linear function of the eye movement features $\boldsymbol{x}_i$ on the viewed documents D_v:

$$p(w_t|\boldsymbol{x}, \boldsymbol{\theta}) = N\left(\sum_{i \in D_v} \boldsymbol{\theta}^T \boldsymbol{x}_i, \sigma^2\right). \quad (3.4)$$

The parameter $\boldsymbol{\theta}$ captures the dependency from the eye movement features to word scores. As before, it is shared across the sessions reflecting our assumption that there exists a common eye movement pattern on relevant words.

The hyperparameter $\boldsymbol{\theta}$ is estimated by maximizing the posterior after marginalizing out the latent query parameter $\boldsymbol{w}$. Document relevances on a new session are given by the predictive distribution $p(r|\boldsymbol{d}, \boldsymbol{x})$, where $\boldsymbol{x}$ are eye movements during the session. In practice, this was approximated by computing a maximum-a-posteriori estimate for the parameter $\boldsymbol{w}$ in (3.4) and substituting that for $\boldsymbol{w}$ in (3.3) to estimate the relevances r.

Experiments

The method was evaluated in experiments that were constructed to simulate a realistic information retrieval scenario. Test subjects first retrieved documents about a given topic using a standard search term-based search engine. They then proceeded to read some of articles returned by the search engine while their eye movements were recorded. After a few articles were read, the model was trained using the eye movements and the relevance labels collected on earlier training sessions and the eye movements on the current session. The learned implicit query was used to re-rank the unread articles. The model's prediction was combined with the original search term-based ranking, and the resulting aggregated ranking was compared with the search term-based ranking. The ranking that used eye movements was clearly better in most sessions.

We also compared the probabilistic model with the SVM-based model from Publication I to test our hypothesis that the training procedure, which avoids the two training stages required by the SVM model, is more efficient. The probabilistic version outperformed the SVM variant.

3.5 Conclusions

This Chapter studied learning an implicit query from eye movement patterns. The learning was done with a machine learning model that uses document relevance feedback from other sessions as an indirect ground truth. In particular, eye movement patterns during reading were used to generalize feedback across sessions. If implicit and explicit feedback has been recorded over long periods of usage, this method can take advantage of the stored feedback to anticipate user's actions even when the user is performing a completely new task, not previously seen on the recorded data. This improves the search performance and the user experience.

The experiments show that it is possible to infer the interest from eye movements at least in a controlled laboratory experimental setting. If eye tracking is cheaply available, it can be combined with explicit relevance judgments or keyword-based search to increase retrieval performance.

One limitation of the experimental results is a small sample size. A larger study would be needed to better evaluate the retrieval performance.

The models presented here are built of standard machine learning components, and the eye movements features were harvested from various psychological studies. Presumably, performance improvements could be achieved by tailoring the models and features to the task.

4. Proactive information access in real world context

This Chapter studies how information related to real world surroundings can be easily accessed. A prototype proactive information retrieval application that merges the retrieved information with the physical environment through an augmented reality display is presented as a potential solution. Eye tracking feedback on the augmented view is collected to decide what to show next. This work was originally published in Publication III.

4.1 Browsing contextual information in physical environments

Publication III presents a novel proactive information retrieval application for accessing information related to the physical or augmented context. The system recognizes nearby people and objects using computer vision techniques. Location is detected by GPS or Bluetooth sensors. User's interests are inferred from interaction patterns captured by a wearable eye tracker, and the retrieval results are blended into a view of the real world using an augmented reality (AR) display.

This essentially turns the physical world into a browsing interface; looking at physical objects or at augmented information is interpreted as relevance feedback, and the augmented information is replaced with newly retrieved, more relevant information. This can be seen as extending the searching by navigation ideas discussed in Section 2.4 into the physical world.

An example use case is a personal assistant that shows contextually related reminders and hints on the wearable display. The displayed information changes according to where the person looking at. If the system notices that one is paying attention to a particular topics, more detailed information about that topic is retrieved. For example, when a person meets someone, the display first shows only a name. If they start a conversation, titles of the latest emails they have exchanged are shown as reminder about what they have discussed before. If the user pays attention to a particular title, more related emails and other documents are opened. Another possible application scenario is a tourist guide, which provides contextually relevant information when visiting new locations.

We have built a prototype of a guide for visitors at a university department and tested the feasibility of the proposed retrieval approach in a small scale user study.

4.2 Related work on mobile information access

Mobile information retrieval (IR) has risen together with prevalence of mobile devices. Mobile IR has certain key differences from retrieval on a desktop computer. Mobile queries tend to be related to spatial and temporal context. According to Sohn et al. [2008], 78% of mobile searches are related to location, current activity, social context or time. Our application makes it particularly easy to access information about the surrounding context just by looking at interesting targets. The interaction capabilities in mobile devices are quite limited, which adds constraints on how the devices are used. The search queries posted from mobile devices tend to be shorter and less complex than queries by desktop users, probably due to restricted text entry facility [Church et al., 2007]. Our approach of inferring feedback from eye movements aims to avoid search terms completely.

During the recent years, there has been an increasing trend to build tools for accessing information related to physical or temporal context. An early example was Jimminy, a system that suggests documents based on physical environment and a note being written [Rhodes, 1997]. Mobile recommender systems suggest relevant data based on feedback from other users and situational factors, such as current location and time [Ricci, 2010, Adomavicius and Tuzhilin, 2011, Pombinho et al., 2012]. Application areas include personalized news recommenders [Yeung and Yang, 2010] and shopping assistants [Sae-Ueng et al., 2008].

Coppola et al. [2010] introduced a contextual mobile web browser that adapts to situations by downloading location-dependent software modules. Thanks to this approach, the browser can customize its inferences based on the environment. Although the general goal of accessing contextually relevant information is similar to ours, we focus more on learning from user feedback.

Many researchers have proposed mobile guide applications for tourists or visitors. For example, the Touring Machine [Feiner et al., 1997] showed virtual tags on top of real university buildings in AR. Van Setten et al. [2004] and Gavalas and Kenteris [2011] both presented tourist guides that recommend interesting nearby locations based on user preferences and recently visited locations. Zimmermann and Lorenz [2008] developed an audio-augmented museum guide that plays sound clips that match to user's location and visit history. A gaze-controlled application for urban exploration by Baldauf et al. [2010] used a speech synthesizer to provide information about viewed buildings. Smartphone applications, such as Layar[1] and Wikitude[2], retrieve georeferenced data from the Internet and show it on a smartphone overlaid on the camera image. These applications can, however, show only static information from a selected data source. By integrating feedback from the eye tracker, we enable applications that react to user's actions.

Truly unobtrusive eye tracker combined with a wearable display will enable developing new kinds of hands-free applications that are aware of user's focus of attention. Many commercial and research solutions exist for separate head-mounted displays and wearable eye trackers, but integrating them into one device seemd to be more demanding. At the moment, a few research groups are building prototype devices. In Publication III, we use a prototype by Nokia Research Center. It uses a novel see-through display technology with a camera for eye tracking [Järvenpää and Aaltonen, 2008]. Park et al. [2008] developed an integrated monocular AR display and eye tracker.

Eye gaze location is a good proxy for the point of attention as was discussed in Section 2.3.3. Therefore, it is an interesting input modality for estimating which of the objects in the current scene are relevant. The most straight-forward way is to assume that the relevance is proportional to the time an object is under visual attention [Qvarfordt and Zhai, 2005]. This, however, may lead to spurious relevance estimates when an object is viewed just to gather more information about it. More advanced estimators learn which features in the gaze pattern are related to relevance and use them for the prediction [Kandemir et al., 2010, Kandemir and Kaski, 2012].

[1] http://layar.com

[2] http://www.wikitude.com

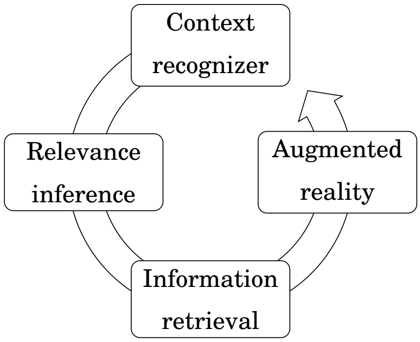

Figure 4.1. Contextual information access loop: Context recognizer detects potential targets of interest. The relevance inference engine uses collected feedback to select the most relevant targets. Contextual information about the selected targets is retrieved and displayed on augmented reality. The augmented information becomes part of the context, and its relevance can be estimated.

4.3 Components of contextual information access system

This Section discusses functionality that is required to implement the proposed information access system. Figure 4.1 shows the main components. The context recognizer module is responsible for identifying faces, objects and other potential targets of interest in the current view. The input for the recognizer module is a video feed from a front-facing camera. The module recognizes faces and barcode markers from the video using online computer vision algorithms. The database of known faces and markers has to be prepared beforehand.

Because not everything visible is equally important, the system must somehow separate relevant and non-relevant objects. It would be distracting to see information related to everything at once. The relevance estimation can be done by inferring preferences from eye movements. If the user does not pay attention to one of the targets of interest detected by the recognizer module, the target's relevance decreases.

The information retrieval module retrieves contextual information about the relevant objects to be displayed on the screen. The retrieval will be done only for the objects that are estimated to be relevant by the relevance inference module. This filtering will help avoid cluttering the screen with useless information.

The retrieved data is displayed on the AR display. The information is rendered next to the corresponding object on the see-through display. If the object moves on the screen, the annotation moves with it. This is a hint to the user that the object and the annotation belong together. The annotations are drawn transparently to avoid occluding other objects.

Because both the physical and augmented contents are displayed on the screen, the relevance of both kind of contents can be estimated by the relevance inference module. In particular, this means that it is possible to learn user's preferences by tracking her reactions to presented content over a period of time. If a certain kind of content selected by the information retrieval module is found relevant, the more content of the same sort should be shown in the future. If, on the other hand, some other kind of content is estimated to be not relevant, the retrieval module can then avoid returning that kind of content. Thus, a student, who has previously shown interest to topics related to studies, will see a list of lectured courses when she meets a professor, whereas a fellow scholar, who shares research interests with the professor, will see titles of the latest publications, when she meets the same professor.

4.4 Pilot application: Virtual laboratory guide

To test the proposed framework, we implemented a prototype application, Virtual Laboratory Guide. It showed information to a visitor about people and objects in a university department. We compared two user interfaces; the first consisted of data glasses with an integrated eye tracker and a scene camera shown in Figure 4.2a [Järvenpää and Aaltonen, 2008], and the second was a tablet computer, where the eye tracking was replaced by explicit pointing.

An intended use case is that a visitor, who arrives at the department, uses the device to see names and other information about the people she meets and about research posters she sees on the corridors. People are recognized using a real-time face recognizer on the camera's video feed, and the posters are detected by 2D bar code markers attached to them. Figure 4.2b shows a screenshot from the application.

Relevance was estimated to be proportional to the time a target was viewed [Qvarfordt and Zhai, 2005]. The information retrieval module fetched short textual annotations for the relevant objects from a database. The database was manually prepared beforehand and consisted of person names, publication titles, office hour information and other similar data. The application could be extended to retrieve data from personal web pages, social media sites, calendars and other public and private sources.

The retrieval module ranked annotations related to relevant objects and

(a) **(b)**

Figure 4.2. (a) Wearable display with an integrated eye tracker. (b) A screenshot from the application showing the real world and augmented information. The white dot shows the eye gaze position. The dot is not visible during normal use.

selected the highest ranking one to be shown on the display. The ranking was based on an estimate of user's interests, which was modeled as a feature vector of topical weights. The dimensions of the feature vector included, among others, user's estimated interest in research and interest in teaching.

The interest vector was continuously updated to reflect which objects and annotations had been estimated to be relevant in the recent past. Each object and annotation had been manually assigned a feature vector that described it. The user's interest vector was periodically updated slightly towards the feature vectors of relevant objects and annotations.

The inference of the user interests enabled the system to adapt the kind of information it showed according to user roles. For students, who had found teaching related information relevant in the past, the system showed information about courses being taught on the department and locations of lecturers' and teaching assistants' offices. For a visiting researcher, who read research related annotations more carefully than teaching related annotations, the guide showed research highlights.

4.4.1 User study

We conducted a usability study to evaluate the effectiveness of the application. Test subjects were asked to wear the device and find answers to given questions. The experiment was held in a room where people and posters were present as information sources that the device recognized. After the experiment, the test subjects answered to a usability questionnaire.

The test subjects reported that the application was helpful in learning new information and they enjoyed using it. When comparing the data glasses to the tablet interface, the users preferred the tablet, because they could look at it when needed, not all the time.

4.5 Conclusions

This Chapter discussed an information access framework for real world environments, originally presented in Publication III. Feasibility of the proposed approach was demonstrated by a pilot application and by a user study.

This work can be seen as a step towards a personal assistant that is able to display contextually relevant notifications at any point. The relevance is inferred from observed interactions. Eye tracking was used here but other sources, for example speech and hand gestures, could be used as well.

Some technical and computational challenges remain. The current integrated wearable displays and eye trackers are still too cumbersome for practical use. It would be interesting to study how the ideas presented in the Chapter could be adapted to smartphone or tablet devices. The eye tracking could perhaps be replaced by a touch interface. More advanced relevance inference methods are needed for improved prediction accuracy and to avoid distracting the user with unnecessary information.

5. Timeline browser with dynamic browsing cues

This Chapter presents the idea of making relevant items more salient to acts as browsing cues and studies if these cues improve retrieval performance. Publications IV and V introduced a retrieval interface that shows images on a timeline and varies their sizes according to estimated relevance. The relevance is estimated from implicit and explicit feedback collected during browsing.

5.1 Dynamic browsing cues for a browser interface

Quite often one is confronted with a situation where one needs to access previously seen digital data objects. Such situations arise, for example, when one needs to re-read a past email or a previously seen web page or review notes from an earlier meeting. Digital material can be captured and stored quite effectively [Gemmell et al., 2006], but searching from a large collection is difficult without proper tools. Publications IV and V propose retrieval interfaces suitable for this task. The interfaces combine search and browsing in a natural way and learn from feedback to suggest relevant items.

The main component in the interface is a timeline that shows images arranged in a temporal order. The images are sized proportionally to their estimated relevance. The relevance estimates change during a search session when more feedback is received. Figure 5.1 shows a screenshot of a dynamic timeline browser prototype.

The motivation for emphasizing relevant images is that it makes them easier to recognize and select. Similar principle was applied in Dasher text entry system [Ward et al., 2000] that allocates screen space to possible textual continuations according to their predicted likelihood. Even if the target image is not among the currently visible ones, the searcher can

(a) Image viewer mode

(b) Result list mode

Figure 5.1. Screenshots from the mobile dynamic timeline browser application from Publication V. A scrollable timeline with images sized according to their estimated relevance values is located at the bottom. The main window can be switched between (a) showing one image in detail and (b) a list of relevant image captions.

use recognized images as browsing cues and decide in which direction to continue browsing. Images are good cues, because humans can recognize a large number of previously seen images [Standing, 1973].

The browsing cues work best when the items can be ordered in a meaningful way so that the user can make justified decisions on which direction to explore next. The temporal order is a good choice, if the user is at least somewhat familiar with the temporal relationships in the dataset. This is the case, for example, for personal data, that is, messages and documents the user has authored or read herself, and for major news events.

The main novelty compared with the earlier browsing cue interfaces discussed in Section 2.4 is that the cues are selected dynamically during the search. The cues change when new feedback is received. The relevance model emphasizes those images that are most likely to be important at the current state of the search. Therefore, they should be easy to recognize. The dynamic changes also create a more browsing-like experience compared with more traditional relevance feedback applications that expect the user to rate one page of items before proceeding. At the same time, the user is still in control of the changes in the interface, because updates only happen in response to explicit feedback.

5.2 Related work on retrieval of personal data

The idea of archiving and searching personal information, such as read books, was introduced already in 1945 by Vannevar Bush [Bush, 1945]. The technology at the time did not, of course, allow actually implementing his vision. Unobtrusive recording of daily activities, such as visited locations, telephone call logs and computer usage, and browsing the collected data with a mobile device started to became feasible in the beginning of 1990's [Lamming and Flynn, 1994]. By 2004, thanks to progress in miniaturization, it was possible to construct small wearable cameras that automatically capture images at frequent intervals, or when the integrated infrared or other sensors detect changes in the environment [Hodges et al., 2011]. Almost all digital content accessed during person's lifetime can now be stored into a unified database that is saved on a standard hard drive [Gemmell et al., 2006]. The only exception is high quality video; storing life-long continuous video takes too much space to be feasible, at least for now.

The usefulness of captured data depends on how easily right artifacts can be summoned up when needed. An effortless access enables retrieval of previously seen items, recollection of episodes in one's life and self-reflection. The remaining of this Section will focus on earlier research on search interfaces for personal data archives.

Timelines are a common way for presenting personal data for browsing. Time is a natural organizing principle, because diverse types of documents can be presented on a unified timeline based on their creation or modification date. A temporal visualization also provides a context in which to interpret the items [Freeman and Gelernter, 1996]. Several researchers have proposed time-based search interfaces. Harada et al. [2004] found that a mobile zoomable timeline-based photo browser performs at least as well as a traditional hierarchy-based interface. Moreover, a hierarchy-based alternative requires additional effort from the user in categorizing the photos. ICLIPS search tool [Chen and Jones, 2010] for personal photo collections combined a timeline and a versatile filtering system. Kim et al. [2006] integrated a timeline view and a map view of personal history. A searcher was able to filter history events by selecting a relevant region on either view. Alonso et al. [2007] proposed showing search results on a timeline instead of the traditional ranked list format. IRemember [Vemuri et al., 2006] used speech recognition to preprocess captured audio and supported a combination of timeline navigation and textual query for retrieving audio files. Annotating the timelines with browsing cues was discussed in Section 2.4.

People remember different kinds of details of the data items in their personal information collections. Therefore, a search system should support searching by rich metadata and allow updating the query interactively

Figure 5.2. The browser interface introduced in Publication IV. Here, the guitar player image is shown at a large size because the mouse cursor is on it. Its neighboring images are also emphasized, but not as much. The two images at the left are relatively large because they have been estimated to be relevant based on received feedback.

after seeing initial results if those bring new associations to one's mind [Cutrell et al., 2006]. Our dynamic timeline interface adapts to evolving information needs by making it possible to give feedback at any stage and by showing the effects of the feedback immediately.

5.3 Experiments and results

We tested the dynamic browser interface in two experiments. The first, which was originally reported in Publication IV, focused on the feasibility of learning the relevance from implicit feedback in a desktop setting. The second experiment was published in Publication V. It studied how well the dynamic browser interface works on a small screen in a mobile device when compared with a more traditional search by textual queries.

5.3.1 Experiment 1: Implicit feedback

This experiment was conducted to find out if the dynamic browser improves retrieval performance, and if the relevance estimates can be learned from implicit feedback. Mouse movements on an image were used as implicit feedback.

A dynamic browser prototype was implemented as a desktop application. It displayed a sequence of images on a timeline. All of the images were visible on the screen all the time. However, because of the large number of images and a limited screen space, the images were shown as tiny thumbnails in their default state. When the mouse cursor was moved on to an image, the image and its neighbors were grown in a fish-eye lens fashion so that they could be viewed in detail. A screenshot of the interface is shown in Figure 5.2.

While the user was moving the mouse cursor over different images to view them, the system recorded the cursor movements among the images.

The total duration the cursor spent on an image and other features were collected as implicit feedback. It was also possible to give explicit feedback by clicking an image to mark it as relevant. A relevance prediction model was learned from feedback collected on training queries, where the ground truth was known.

The image sizes were modulated by two factors: by the relevance prediction that was cumulative over the whole search session and, temporarily, by the fish-eye zooming, when the mouse cursor was hovering over an image.

Because collecting real personal data from every test subject for the experiments would have been too laborious and would have complicated the intrasubject analysis, we simulated personal history by asking all the test subjects to view the same movie beforehand. The movie can be though as becoming part of test subjects' histories in the sense that they were familiar with its content.

The test subjects were asked to complete six recall tasks using our interface. The images shown on the interface were snapshots from the movie. The tasks involved finding images that contained certain sets of people or objects.

The test subjects were able to find target images faster with the dynamic browser than with an alternative interface that did not modify the image sizes according to the relevance prediction. The relevance predictions of a model that combines implicit and explicit feedback were slightly better than using explicit feedback alone. The result shows promise in the idea of modulating the image size with the estimated relevance.

5.3.2 Experiment 2: Mobile interface

The second experiment studied how well the dynamic browser works on a mobile device with a constrained screen space. If one can access the history on a mobile device, the browser is more likely to be available when a need arises. Mobility is also important for recording the data.

On the general level the interface of the mobile dynamic browser is similar to the desktop version in the first experiment; the images are shown on the timeline and their sizes are changed dynamically when new feedback is received. However, because a mobile device has very limited screen space, it is not possible to fit even small thumbnails of every image on the timeline. Therefore, only a small part of the timeline is visible, and the rest is available by scrolling. The locations of relevant images, which are not currently in the view, are shown as small dots on the scrollbar. Screenshots of this version of the application are shown in Figure 5.1.

The data used in the experiments was from daily TV news broadcasts. Each news event was shown in the interface as a key frame and a short textual description of the event. The news are a form of "shared history," because the test subjects are likely to remember at least the major news events.

In the experiment, it was possible to search the news events with a textual query or by browsing the timeline. The relevance prediction combined relevance feedback on visual and textual feature spaces and the search results. Only explicit feedback was collected in this experiment.

The search performance was better with the dynamic browser than with the more traditional textual search; the test subjects were able to find images related to given tasks with less effort. We also wanted to known how much does the temporal ordering of the images contribute to the result. To measure this, we compared the dynamic timeline interface with a similar interface, where the temporal ordering of the images was replaced by relevance ordering. The timeline interface was more efficient in locating the relevant news events, which shows that the cues were more useful when the ordering was (at least partly) familiar to the subjects.

5.4 Conclusions

This Chapter introduced a search interface that employs browsing cues by modulating the sizes of the data items according to their relevance. The relevance estimates were updated online based on the feedback collected during the search session.

Two prototypes of the interface were tested in two user studies. The experiments showed that it takes less effort to find correct items with the proposed dynamic browser than with more traditional search interfaces both in desktop and in mobile environments.

Although the experiments showed that the dynamic browser interface improves retrieval performance, it is not quite clear how much the familiarity with the data affects one's performance. The hypothesis is that familiarity with the data should increase subject's ability to take advantage of the browsing cues. It would also be interesting to collect even anecdotal evidence how well the interface works on real personal history data. Another interesting direction of further study would be to predict not just relevance in current session but also other aspects that might make items good browsing cues. For example, some items are more likely to be remembered than others, because they provoke more meaningful associations [Horvitz et al., 2004]. Furthermore, modeling the exploration/exploitation trade-off when selecting the cues might improve prediction performance [Auer et al., 2010, Glowacka et al., 2013].

6. Inference of relevance with multiple data domains

This Chapter studies the problem of combining relevance feedback given to connected data domains, for example images and their captions. A relevance inference method that combines feedback across domains is introduced. Inferences can be done also when the domains are only partially connected, that is, co-occurring text is known only for some images. The method was originally introduced in Publication VI.

6.1 Problem setting

An example scenario where one might want to give feedback on multiple types of data is retrieval from personal history. People tend to remember different details of the items they are looking for. Therefore, a good retrieval system should make it possible to give feedback on several types of data and retrieve results consistent with all feedback.

A person who is looking for notes from a particular meeting might recall only the general topic of the meeting and a diagram from slides shown in the meeting. A flexible search interface should allow giving feedback for text documents that discuss topics relevant for the meeting, and for images that resemble the diagram. If the search engine can integrate both feedback sources, it should learn to return results from the correct meeting with both the topic and the image content matching the received feedback.

Combining feedback from different data sources is possible only if there exist some suitable connections between the domains. Typically these connections come as co-occurrences. The data items on different domains are connected if they co-occur in time, were recorded at the same location or share some relevant property. In the meeting retrieval example above, the meeting notes and the diagram were connected in the database, because

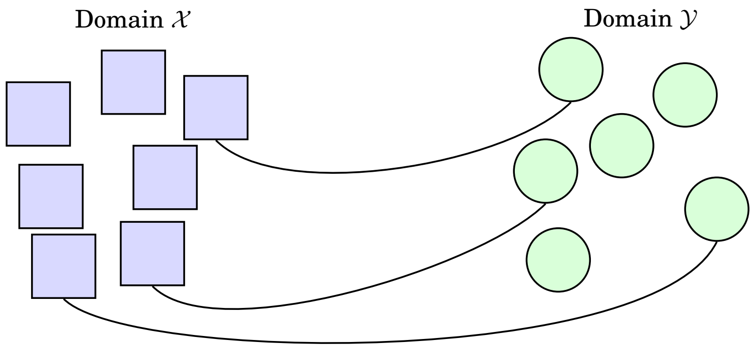

Figure 6.1. The method combines relevance feedback that has been given on multiple data domains. The domains are assumed to be connected through co-occurring items. The connections, represented here as lines, are known only for a subset of items.

they were captured during the same calendar event. Other examples of co-occurrences are images and text that are located on the same web page. The retrieval method presented here does not make any assumptions on what kind of connections are being used. It is simply assumed that a database with connected items is available.

The method discussed in this Chapter does not assume that all items are connected between the domains. In fact, the method is meant for a case where only a small fraction of items have known co-occurrences as shown in Figure 6.1. The most of the items have a representation on one data domain only.

The search interface should show different data domains and make it possible to give feedback on any domain. An example could be showing top 10 most relevant items as lists for each domain separately. The display should be updated when feedback is received. Publication VI focuses on the ranking method and does not consider the functionality of the interface further.

6.2 Multi-view learning

Combining evidence from multiple sources is an active research area in machine learning. This Section gives a short review of earlier work.

The multi-view learning studies learning when several representations of items are available. The representations correspond to the connected items on different data domains in our formulation. Jointly modeling all available data is expected to be beneficial. Application areas include modeling relationships between images and captions [Blei and Jordan, 2003], learning common semantics from translated documents [Vinokourov et al., 2003] and integrating mRNA and protein expression measurements in bioinformatics [Rogers et al., 2010], among others.

Multiple views can act as regularization in a semi-supervised task where only some items are labeled. Co-training [Blum and Mitchell, 1998] is an iterative bootstrapping-like algorithm, which alternates between learning view-specific classifiers from the labeled samples and generating pseudo-labels for the unlabeled items when at least one view-specific classifier is confident on its prediction. Co-regularization [Brefeld et al., 2006, Farquhar et al., 2006, Yu et al., 2011] regularizes learning of view-specific classifiers by imposing a penalty term on the cost function if the predictions learned from different views disagree.

Our work differs from most multi-view learning, because we assume that no one-to-one correspondence between the data domains exists, that is, most items have missing views. Therefore, the standard multi-view learning methods are not applicable.

6.3 Learning rankers by transferring feedback

This Section describes a relevance inference method that learns a ranking function by combining direct relevance feedback for a domain and feedback transferred from other domains. The method works by learning a ranking function for each data domain with a constraint that the ranking functions should be consistent over the connected items.

We assume the binary feedback is given for the domains separately. An item on a domain is labeled either as relevant or non-relevant or is left unlabeled.

Separate ranking algorithms are learned for each domain. We formulate the problem so that any ranking algorithm that utilizes binary labels can be used. For the experiments, we choose the SVM-rank algorithm [Joachims, 2002].

To ensure that the domain-specific ranking functions are consistent over the different domains, we transfer some information between the domains. Specifically, we construct *pseudo labels* for some of the unlabeled items. The pseudo labels are selected in a way that drives the rankings to be more consistent with each other.

Feedback is transferred from the source domain, denoted by $\mathcal{Y}$, to the target domain, denoted by $\mathcal{X}$. Note that the identities of the domains are interchangeable. For example, if the two domains are texts and images, feedback can be transferred either from text to images or from images to texts.

The complete algorithm for transferring feedback and learning a ranking function for a domain is composed of the following steps:

1. Learn a ranking for the source domain $\mathcal{Y}$ using the chosen single-domain ranking algorithm.

2. Transfer the feedback from $\mathcal{Y}$ to $\mathcal{X}$ using pseudo labels as outlined below.

3. Learn a single-domain ranking for the target domain $\mathcal{X}$ from the feedback given directly to $\mathcal{X}$ and the feedback transferred from $\mathcal{Y}$ in step 2.

The first and the third steps use standard single-domain ranking function learning algorithms. The second step is more interesting. It is discussed below.

Figure 6.2 illustrates the feedback transfer procedure. For each unlabeled item $\boldsymbol{x}_i \in \mathcal{X}$, which is connected to an item $\boldsymbol{y}_i$ on the other domain, we check if there exists a labeled item on $\mathcal{Y}$ that is similar to $\boldsymbol{y}_i$. If such item exists, a pseudo label for $\boldsymbol{x}_i$ is set to the same value as the found label. The similarity is measured in rank position sense; if two items are at most K positions apart in current ranking of $\mathcal{Y}$, they are considered similar. If $\mathcal{Y}$ contains more than one suitable item, the pseudo label is chosen by a majority vote.

6.4 Experiments on ranking images and captions

In Publication VI, the algorithm was tested on a dataset with images and their textual captions as the two domains. Only 5% of the image-caption pairs were known to the algorithm, the rest were hidden. A set of 40 queries with manually constructed ground truth labels were used to simulate feedback. Positive and negative samples were randomly chosen and used as simulated relevance feedback to train the algorithm. We ran several tests transferring feedback from images to text and the other way

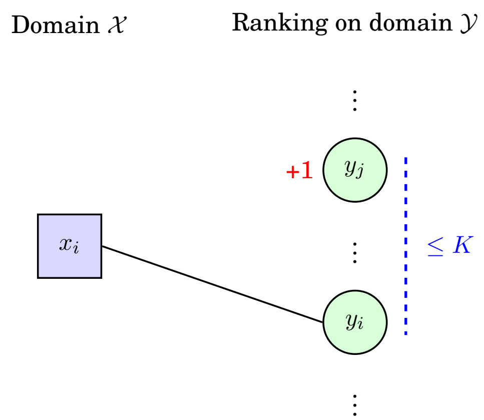

Figure 6.2. An example pseudo label construction process for an initially unlabeled item x_i. A ranking has been learned on domain $\mathcal{Y}$ from the feedback (shown in red) given by the user. The pseudo label of $\boldsymbol{x}_i$ will be set to relevant $(+1)$ because the connected item, $\boldsymbol{y}_i$, on the other domain is no farther than K rank positions from an item $\boldsymbol{y}_j$ that has been explicitly labeled by the user as relevant.

around with different amounts of simulated feedback.

We evaluated the learned rankings against the ground truth label and observed that the ranking learned from the combination of domain-specific and transferred feedback was better than ranking learned from the domain-specific feedback only. This shows that it is possible to use relevance feedback on other related data domains to improve relevance inference.

6.5 Conclusions

This Chapter presented a relevance inference method for cases when feedback is available on multiple domains. The algorithm integrates feedback from other domains to learn an improved ranking function. The proposed algorithm is meant especially for a case where only some connections between the domains are known. This Chapter together with Chapter 3 shows that relevance estimates can often be improved by combining evidence from multiple sources with proper modeling.

So far the method has been evaluated only in an artificial experiment. It should be tested on a more realistic task. The feedback transferring method is largely heuristic. A possible future direction would be formulating a more principled way of constraining the domain-specific rankings to be similar with each other.

7. Conclusions

This thesis has discussed intelligent information access methods and new kinds of retrieval applications made possible by modeling user's intent. Better search systems are important because they enable faster and more effortless access to information. The importance of search tools as an interface to information is likely to increase in the future as the amount of digital data continues to grow at a rapid pace. Furthermore, innovations in search technology will facilitate completely new kinds of applications, such as the real world information access tool introduced in this thesis.

Information retrieval (IR) can benefit from accurate user modeling in several ways. This thesis focused on two directions: proactive information retrieval and browsing cues. The thesis introduced machine learning methods that infer search terms from reading patterns (Publications I and II). The novelty here is that the relevance is inferred from data unlike other works that rely on hand-built rules [Buscher et al., 2008]. Information access in the context of the physical environment was introduced in Publication III as a new application area for proactive retrieval. Typically proactive retrieval is employed in desktop applications, where collecting feedback is relatively easy. The work here reaches outside a desktop setting with the help of wearable eye trackers that record user's interaction with both physical objects and augmented information.

Earlier research [Ringel et al., 2003, Lee et al., 2008] indicates that static browsing cues improve retrieval performance, because they aid in understanding the context of what is shown on the display. This thesis extended those previous works by employing dynamic cues that adapt to user's goals (Publications IV and V). The adaption was done by collecting relevance feedback during the search session. Dynamic cues are useful, because they are selected according to user's actions and should therefore be easily recognizable to her.

Machine learning methods for relevance inference were presented in Publications I, II and VI. The methods improve relevance prediction on one search session by bringing in evidence from related search sessions. With appropriate modeling, the relevance inferences can be transferred even if the search goals of the sessions are different. This kind of approach will result in a smoother retrieval experience because more reliable inferences can be made with smaller amount of direct feedback.

The empirical user studies presented here show that intelligent information retrieval is feasible in many different kind of applications. These findings complement earlier research on implicit feedback (e.g. [Puolamäki et al., 2005, Joachims and Radlinski, 2007]) and proactive retrieval [Rhodes, 2000, Laqua et al., 2011] by extending into new application areas. While many interesting retrieval concepts have been presented by this thesis and by others, only a few have matured into everyday applications. Researchers and search engine developers should work together to bring advances in IR research into common use.

The user studies conducted in this thesis were rather small scale experiments in laboratory conditions. More comprehensive studies are needed to ensure the reliability of the findings in real life conditions. One issue requiring careful analysis is that feedback for search is likely to be even noisier, because the search is not user's main focus, instead the functionality of the search is to support the main task.

Eye tracking is a promising source of implicit feedback, because gaze is a natural interaction modality. However, technological and economical challenges have so far prevented mass production and greater adoption of the technology. The researchers can drive the demand for eye tracking technology by continuing exploring the possibilities of eye tracking interfaces. At the same time, the research on other implicit feedback channels, such as interaction patterns and hand gestures, is needed to evaluate their merits. The best results can most likely be achieved by studying machine learning methods for combining multiple implicit and explicit feedback modalities.

The proactive eye tracking based retrieval, the contextual information access tool and the personal history browser are three examples of applications made possible by a combination of predicting user's needs and intelligent retrieval. Exploring also other novel ways of making relevant information easier to access at the right time continues to provide plenty of opportunities for research in the future.

An interesting line of future research are personal assistants that are able to show notifications related to the current context. They infer interests from the interaction with the recommendations and the environment. The interactions can be recorded by wearable eye trackers or by other means. The inference methods and browsing interfaces presented in this thesis are steps towards such assistants. Open questions not answered in here include, among others, what kind of information can and should be displayed by such assistants and how the information should be presented to notify the user about relevant events but, at the same time, avoid distracting the user with unnecessary clutter.

This is an interesting time in information retrieval research. On the one hand, more and more digital data becomes available all the time stressing the need for practical search tools, on the other hand, mobility, eye tracking and other new technologies can provide search engines with much more information for inferring user's goals. Together the two trends imply a plenty of opportunities for employing machine learning methods in intelligent search systems.

Bibliography

I. J. Aalbersberg. Incremental relevance feedback. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 11–22. ACM, New York, NY, 1992.

G. Adomavicius and A. Tuzhilin. *Recommender systems handbook*, chapter Context-Aware Recommender Systems, pages 217–253. Springer, Berlin, Germany, 2011.

E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 19–26. ACM, New York, NY, 2006.

J. S. Agustin, H. Skovsgaard, E. Mollenbach, M. Barret, M. Tall, D. W. Hansen, and J. P. Hansen. Evaluation of a low-cost open-source gaze tracker. In *Proceedings of the Symposium on Eye-Tracking Research & Applications (ETRA)*, pages 77–80. ACM, New York, NY, 2010.

O. Alonso, R. Baeza-Yates, and M. Gertz. Exploratory search using timelines. In *SIGCHI Workshop on Exploratory Search and HCI Workshop*, pages 23–26. Microsoft Research, Redmond, WA, 2007.

P. Auer, Z. Hussain, S. Kaski, A. Klami, J. Kujala, J. Laaksonen, A. P. Leung, K. Pasupa, and J. Shawe-Taylor. Pinview: Implicit feedback in content-based image retrieval. In *Proceedings of Workshop on Applications of Pattern Analysis*, volume 11 of *JMLR Workshop and Conference Proceedings*, pages 51–57, 2010.

B. P. Bailey and S. T. Iqbal. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction*, 14(4):21:1–21:28, 2008.

M. Baldauf, P. Fröhlich, and S. Hutter. KIBITZER: a wearable system for eye-gaze-based mobile urban exploration. In *Proceedings of the 1st Augmented Human International Conference*, pages 9:1–9:5. ACM, New York, NY, 2010.

D. Barreau and B. A. Nardi. Finding and reminding: file organization from the desktop. *ACM SIGCHI Bulletin*, 27(3):39–43, 1995.

M. J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5):407–424, 1989.

R. Bednarik, H. Vrzakova, and M. Hradis. What do you want to do next: a novel approach for intent prediction in gaze-based interaction. In *Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA)*, pages 83–90. ACM, New York, NY, 2012.

N. J. Belkin. Helping people find what they don't know. *Communications of the ACM*, 43(8):58–61, 2000.

O. Bergman, R. Beyth-Marom, R. Nachmias, N. Gradovitch, and S. Whittaker. Improved search engines and navigation preference in personal information management. *ACM Transactions on Information Systems*, 26(4):20:1–20:24, 2008.

O. Bergman, M. Tene-Rubinstein, and J. Shalom. The use of attention resources in navigation versus search. *Personal and Ubiquitous Computing*, pages 1–8, 2012.

R. Biedert, G. Buscher, and A. Dengel. The eyeBook – using eye tracking to enhance the reading experience. *Informatik-Spektrum*, 33(3):272–281, 2010.

D. Bilal. Children's use of the Yahooligans! web search engine: I. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 51(7):646–665, 2000.

D. Billsus, D. M. Hilbert, and D. Maynes-Aminzade. Improving proactive information systems. In *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI)*, pages 159–166. ACM, New York, NY, 2005.

D. M. Blei and M. I. Jordan. Modeling annotated data. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 127–134. ACM, New York, NY, 2003.

A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT)*, pages 92–100. ACM, New York, NY, 1998.

R. Boardman and M. A. Sasse. "Stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 583–590. ACM, New York, NY, 2004.

M. Bonfils. Yandex just passed Bing to become 4th largest global search engine, 2013. URL `http://searchenginewatch.com/article/2242374/Yandex-Just-Passed-Bing-to-Become-4th-Largest-Global-Search-Engine`. Retrieved Feb 13, 2013.

U. Brefeld, T. Gärtner, T. Scheffer, and S. Wrobel. Efficient co-regularised least squares regression. In *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, pages 137–144. ACM, New York, NY, 2006.

G. Buscher, A. Dengel, and L. van Elst. Query expansion using gaze-based feedback on the subdocument level. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 387–394. ACM, New York, NY, 2008.

G. Buscher, A. Dengel, R. Biedert, and L. van Elst. Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. *ACM Transactions on Interactive Intelligent Systems*, 1(2):9:1–9:30, 2012.

V. Bush. As we may think. *The Atlantic Monthly*, 176:101–108, July 1945.

I. Campbell. Interactive evaluation of the ostensive model using a new test collection of images with multiple relevance assessments. *Information Retrieval*, 2(1):87–114, 2000.

R. Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, 1997.

Y. Chen and G. J. F. Jones. Augmenting human memory using personal lifelogs. In *Proceedings of the First Augmented Human International Conference*, pages 24:1–24:9. ACM, New York, NY, 2010.

K. Church, B. Smyth, P. Cotter, and K. Bradley. Mobile information access: A study of emerging search behavior on the mobile internet. *ACM Transactions on the Web*, 1(1):4:1–4:38, 2007.

P. Coppola, V. D. Mea, L. D. Gaspero, D. Menegon, D. Mischis, S. Mizzaro, I. Scagnetto, and L. Vassena. The context-aware browser. *IEEE Intelligent Systems*, 25(1):38–47, 2010.

C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3): 273–297, 1995.

E. Cutrell, D. Robbins, S. Dumais, and R. Sarin. Fast, flexible filtering with Phlat. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 261–270. ACM, New York, NY, 2006.

M. Czerwinski, S. Dumais, G. Robertson, S. Dziadosz, S. Tiernan, and M. van Dantzich. Visualizing implicit queries for information management and retrieval. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 560–567. ACM, New York, NY, 1999.

C. Dhanjal, S. R. Gunn, and J. Shawe-Taylor. Sparse feature extraction using generalised partial least squares. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 27–32. IEEE, Los Alamitos, CA, 2006.

A. R. Doherty and A. F. Smeaton. Automatically augmenting lifelog events using pervasively generated content from millions of people. *Sensors*, 10(3), 2010.

A. T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag, London, UK, 2nd edition, 2007.

G. B. Duggan and S. J. Payne. Skim reading by satisficing: evidence from eye tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1141–1150. ACM, New York, NY, 2011.

S. Dumais, E. Cutrell, R. Sarin, and E. Horvitz. Implicit queries (IQ) for contextualized search. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 594. ACM, New York, NY, 2004.

D. Elsweiler, I. Ruthven, and C. Jones. Towards memory supporting personal information management tools. *Journal of the American Society for Information Science and Technology*, 58(7):924–946, 2007.

J. D. R. Farquhar, D. Hardoon, H. Meng, J. Shawe-Taylor, and S. Szedmak. Two view learning: SVM-2K, theory and practice. In *Advances in Neural Information Processing Systems (NIPS) 18*, pages 355–362. MIT Press, Cambridge, MA, 2006.

S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A Touring Machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Personal and Ubiquitous Computing*, 1(4):208–217, 1997.

M. Fejtová, J. Fejt, and L. Lhotská. Controlling a PC by eye movements: The MEMREC project. In *Computers Helping People With Special Needs*, pages 770–773. Springer, Berlin, Germany, 2004.

S. Fox, K. Karnawat, M. Mydland, S. Dumais, and T. White. Evaluating implicit measures to improve web search. *ACM Transactions on Information Systems*, 23(2):147–168, 2005.

E. Freeman and D. Gelernter. Lifestreams: a storage model for personal data. *ACM SIGMOD Record*, 25(1):80–86, 1996.

M. Fuller, L. Kelly, and G. J. Jones. Applying contextual memory cues for retrieval from personal information archives. In *Personal Information Management in Conjunction with CHI 2008 Workshop*. Microsoft Research, Redmond, WA, 2008.

D. Gavalas and M. Kenteris. A web-based pervasive recommendation system for mobile tourist guides. *Personal and Ubiquitous Computing*, 15(7):759–770, 2011.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL, 2nd edition, 2004.

J. Gemmell, G. Bell, and R. Lueder. MyLifeBits: a personal database for everything. *Communications of the ACM*, 49(1):88–95, 2006.

C. Giraud-Carrier, R. Vilalta, and P. Brazdil. Special issue on meta-learning. *Machine Learning*, 54(3):187–312, 2004.

D. Glowacka, T. Ruotsalo, K. Konuyshkova, K. Athukorala, S. Kaski, and G. Jacucci. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proceedings of the 18th International Conference on Intelligent User Interface (IUI)*, pages 117–128. ACM, New York, NY, 2013.

S. Grandhi and Q. Jones. Technology-mediated interruption management. *International Journal of Human-Computer Studies*, 68(5):288–306, 2010.

K. Gyllstrom and C. Soules. Seeing is retrieving: building information context from what the user sees. In *Proceedings of the 13th International Conference on Intelligent User Interfaces (IUI)*, pages 189–198. ACM, New York, NY, 2008.

J. Hailpern, N. Jitkoff, A. Warr, K. Karahalios, R. Sesek, and N. Shkrob. YouPivot: improving recall with contextual search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1521–1530. ACM, New York, NY, 2011.

J. P. Hansen, D. W. Hansen, and A. S. Johansen. Bringing gaze-based interaction back to basics. In *Universal Access In HCI: Towards an Information Society for All*, pages 325–328. Lawrence Erlbaum Associates, Mahwah, NJ, 2001.

S. Harada, M. Naaman, Y. J. Song, Q. Wang, and A. Paepcke. Lost in memories: interacting with photo collections on PDAs. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 325–333. ACM, New York, NY, 2004.

D. R. Hardoon and K. Pasupa. Image ranking with implicit feedback from eye movements. In *Proceedings of the Symposium on Eye-Tracking Research & Applications (ETRA)*, pages 291–298. ACM, New York, NY, 2010.

E. Hargittai. Classifying and coding online actions. *Social Science Computer Review*, 22(2):210–227, 2004.

S. Hodges, E. Berry, and K. Wood. SenseCam: A wearable camera that stimulates and rehabilitates autobiographical memory. *Memory*, 19(7):685–696, 2011.

F. Hopfgartner, D. Hannah, and N. G. an Joemon M. Jose. Capturing multiple interests in news video retrieval by incorporating the ostensive model. In *2nd International Workshop on Personalized Access, Profile Management and Context Awareness: Databases*. IMIS, Athens, Greece, 2008.

E. Horvitz, S. Dumais, and P. Koch. Learning predictive models of memory landmarks. In *26th Annual Meeting of the Cognitive Science Society*, pages 583–588. Lawrence Erlbaum Associates, Mahwah, NJ, 2004.

K.-S. Hwang and S.-B. Cho. Landmark detection from mobile life log using a modular Bayesian network model. *Expert Systems with Applications*, 36(10): 12065–12076, 2009.

A. Hyrskykari, P. Majaranta, and K.-J. Räihä. Proactive response to eye movements. In *Proceedings of 9th IFIP INTERACT Conference on Human-Computer Interaction*, pages 129–136. IOS Press, Amsterdam, The Netherlands, 2003.

Y. Ishiguro, A. Mujibiya, T. Miyaki, and J. Rekimoto. Aided eyes: eye activity sensing for daily life. In *Proceedings of the 1st Augmented Human International Conference*, pages 25:1–25:7. ACM, New York, NY, 2010.

T. Järvenpää and V. Aaltonen. *Photonics in Multimedia II*, volume 7001 of *Proceedings of SPIE*, chapter Compact near-to-eye display with integrated gaze tracker, pages 700106:1–700106:8. SPIE, 2008.

T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 133–142. ACM, New York, NY, 2002.

T. Joachims and F. Radlinski. Search engines that learn from implicit feedback. *IEEE Computer*, 40(8):34–40, 2007.

G. J. F. Jones and P. J. Brown. Context-aware retrieval for ubiquitous computing environments. In *Mobile and Ubiquitous Information Access*, pages 227–243. Springer-Verlag, Berlin, Germany, 2004.

S. Kandel, A. Paepcke, J. Hellerstein, and J. Heer. Wrangler: interactive visual specification of data transformation scripts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3363–3372. ACM, New York, NY, 2011.

M. Kandemir and S. Kaski. Learning relevance from natural eye movements in pervasive interfaces. In *Proceedings of the International Conference on Multimodal Interaction (ICMI)*, pages 85–92. ACM, New York, NY, 2012.

M. Kandemir, V.-M. Saarinen, and S. Kaski. Inferring object relevance from gaze in dynamic scenes. In *Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA)*, pages 105–108. ACM, New York, NY, 2010.

D. Kelly and N. J. Belkin. Display time as implicit feedback: understanding task effects. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 377–384. ACM, New York, NY, 2004.

D. Kelly and J. Teevan. Implicit feedback for inferring user preference: a bibliography. *ACM SIGIR Forum*, 37(2):18–28, 2003.

I.-J. Kim, S. C. Ahn, H. Ko, and H. Kim. PERSONE: personalized experience recoding and searching on networked environment. In *Proceedings of the 3rd ACM Workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE)*, pages 49–54. ACM, New York, NY, 2006.

A. Klami, C. Saunders, T. de Campos, and S. Kaski. Can relevance of images be inferred from eye movements? In *Proceeding of the 1st ACM International Conference on Multimedia Information Retrieval (MIR)*, pages 134–140. ACM, 2008.

A. Kobsa. *The Adaptive Web: Methods and Strategies of Web Personalization*, chapter Generic User Modeling Systems, pages 136–154. Springer, Berlin, Germany, 2007.

E. Kowler. Eye movements: The past 25 years. *Vision Research*, 51(13):1457–1483, 2011.

L. Kozma, A. Klami, and S. Kaski. GaZIR: Gaze-based zooming interface for image retrieval. In *Proceedings of The Eleventh International Conference on Multimodal Interfaces and The Sixth Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI)*, pages 305–312. ACM, New York, NY, 2009.

R. Kraft, C. C. Chang, F. Maghoul, and R. Kumar. Searching with context. In *Proceedings of the 15th International Conference on World Wide Web*, pages 477–486. ACM, New York, NY, 2006.

M. Lamming and M. Flynn. Forget-me-not: Intimate computing in support of human memory. In *Proceedings of FRIEND21 International Symposium on Next Generation Human Interfaces*, pages 125–128. Rank Xerox, Norwalk, CT, 1994.

S. Laqua, M. A. Sasse, S. Greenspan, and C. Gates. Do you know dis?: a user study of a knowledge discovery tool for organizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2887–2896. ACM, New York, NY, 2011.

S. Lawrence. Context in web search. *IEEE Data Engineering Bulletin*, 23(3): 25–32, 2000.

H. Lee, A. F. Smeaton, N. O'Connor, G. Jones, M. Blighe, D. Byrne, A. Doherty, and C. Gurrin. Constructing a SenseCam visual diary as a media process. *Multimedia Systems Journal*, 14(6):341–349, 2008.

S.-H. Lee, J.-Y. Lee, and J.-S. Choi. Design and implementation of an interactive HMD for wearable AR system. In *17th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)*, pages 1–6. IEEE, Los Alamitos, CA, 2011.

T. D. Loboda, P. Brusilovsky, and J. Brunstein. Inferring word relevance from eye-movements of readers. In *Proceedings of the 16th International Conference on Intelligent User Interfaces (IUI)*, pages 175–184. ACM, New York, NY, 2011.

P. P. Maglio and C. S. Campbell. Attentive agents. *Communications of the ACM*, 46:47–51, 2003.

C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, Cambridge, UK, 2008.

R. Mantiuk, M. Kowalik, A. Nowosielski, and B. Bazyluk. Do-it-yourself eye tracker: Low-cost pupil-based eye tracker for computer graphics applications. In *Advances in Multimedia Modeling*, pages 115–125. Springer, Berlin, Germany, 2012.

T. Miller and S. Agne. Attention-based information retrieval using eye tracker data. In *Proceedings of the 3rd International Conference on Knowledge Capture*, pages 209–210. ACM, New York, NY, 2005.

M. Morita and Y. Shinoda. Information filtering based on user behavior analysis and best match text retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 272–281. Springer-Verlag, Berlin, Germany, 1994.

S. Nilsson, T. Gustafsson, and P. Carleberg. Hands free interaction with virtual information in a real environment: Eye gaze as an interaction tool in an augmented reality system. *Psychnology Journal*, 7(2):175—-196, 2009.

T. Ohno. EyePrint: support of document browsing with eye gaze trace. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, pages 16–23. ACM, New York, NY, 2004.

O. Oyekoya and F. Stentiford. Perceptual image retrieval using eye movements. In *International Workshop on Intelligent Computing in Pattern Analysis/Synthesis*, Advances in Machine Vision, Image Processing, and Pattern Analysis, pages 281–289. Springer, Berlin, Germany, 2006.

H. M. Park, S. H. Lee, and J. S. Choi. Wearable augmented reality system using gaze interaction. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 175–176. IEEE, Los Alamitos, CA, 2008.

T. Pfeiffer, M. E. Latoschik, and I. Wachsmuth. Evaluation of binocular eye trackers and algorithms for 3D gaze interaction in virtual reality environments. *Journal of Virtual Reality and Broadcasting*, 5(16), 2008.

P. Pombinho, M. B. Carmo, and A. P. Afonso. Context aware point of interest adaptive recommendation. In *Proceedings of the 2nd Workshop on Context-awareness in Retrieval and Recommendation (CARR)*, pages 30–33. ACM, New York, NY, 2012.

M. Pomplun and S. Sunkara. Pupil dilation as an indicator of cognitive workload in human-computer interaction. In *Proceedings of the International Conference on HCI*, pages 542–546. Lawrence Erlbaum Associates, Mahwah, NJ, 2003.

K. Puolamäki, J. Salojärvi, E. Savia, J. Simola, and S. Kaski. Combining eye movements and collaborative filtering for proactive information retrieval. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 146–153. ACM, New York, NY, 2005.

P. Qvarfordt and S. Zhai. Conversing with the user based on eye-gaze patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 221–230. ACM, New York, NY, 2005.

O. M. Ramos, M. C. P. Melguizo, L. Boves, and M. G. C. Vergara. Optimal access to information while writing: writing in the internet age. In *Proceedings of the Second International Symposium on Information Interaction in Context*, pages 139–144. ACM, New York, NY, 2008.

K. Rayner. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8):1457–1506, 2009.

B. Rhodes. The wearable remembrance agent: A system for augmented memory. *Personal Technologies Journal Special Issue on Wearable Computing*, 1:218–224, 1997.

B. Rhodes and T. Starner. Remembrance agent: A continuously running automated information retrieval system. In *Proceedings of The First International Conference on The Practical Application of Intelligent Agents and Multi Agent Technology*, pages 122–125. AAAI, Palo Alto, CA, 1996.

B. J. Rhodes. Margin notes: building a contextually aware associative memory. In *Proceedings of the 5th International Conference on Intelligent User Interfaces (IUI)*, pages 219–224. ACM, New York, NY, 2000.

F. Ricci. Mobile recommender systems. *Information Technology and Tourism*, 12 (3):205–231, 2010.

M. Ringel, E. Cutrell, S. Dumais, and E. Horvitz. Milestones in time: The value of landmarks in retrieving information from personal stores. In *Proceedings of the 9th IFIP INTERACT Conference on Human-Computer Interaction*, pages 184–191. IOS Press, Amsterdam, The Netherlands, 2003.

S. Rogers, A. Klami, J. Sinkkonen, M. Girolami, and S. Kaski. Infinite factorization of multiple non-parametric views. *Machine Learning*, 79(1–2):201–226, 2010.

I. Ruthven, A. Tombros, and J. Jose. A study on the use of summaries and summary-based query expansion for a question-answering task. In *23rd BCS European Annual Colloquium on Information Retrieval Research*, pages 41–53. BCS, London, UK, 2001.

S. Sae-Ueng, S. Pinyapong, A. Ogino, and T. Kato. Personalized shopping assistance service at ubiquitous shop space. In *22nd International Conference on Advanced Information Networking and Applications (AINAW) – Workshops*, pages 838–843. IEEE, Los Alamitos, CA, 2008.

J. Salojärvi, I. Kojo, J. Simola, and S. Kaski. Can relevance be inferred from eye movements in information retrieval? In *Proceedings of Workshop on Self-Organizing Maps (WSOM)*, pages 261–266. Kyushu Institute of Technology, Fukuoka, Japan, 2003.

J. Salojärvi, K. Puolamäki, and S. Kaski. Implicit relevance feedback from eye movements. In *Artificial Neural Networks: Biological Inspirations — ICANN 2005*, volume I, pages 513–518. Springer, Berlin, Germany, 2005.

G. Salton, editor. *The SMART retrieval system: Experiments in automatic document processing*. Prentice-Hall, Englewood Cliffs, NJ, 1971.

G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. *Journal of the American society for information science*, 41(4):288–297, 1990.

J. Simola, J. Salojärvi, and I. Kojo. Using hidden markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research*, 9(4):237—251, 2008.

T. Sohn, K. A. Li, W. G. Griswold, and J. D. Hollan. A diary study of mobile information needs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 433–442. ACM, New York, NY, 2008.

C. A. N. Soules and G. R. Ganger. Connections: using context to enhance file search. In *Proceedings of the 20th ACM Symposium on Operating Systems Principles*, pages 119–132. ACM, New York, NY, 2005.

A. H. Spink, H. Greisdorf, and J. Bateman. From highly relevant to not relevant: examining different regions of relevance. *Information Processing and Management*, 34(5):599–622, 1998.

SRI International. CALO: Cognitive assistant that learns and organizes. URL `http://www.ai.sri.com/project/CALO`. Retrieved Jan 30, 2013.

L. Standing. Learning 10,000 pictures. *The Quarterly Journal of Experimental Psychology*, 25:207–222, 1973.

I. Starker and R. A. Bolt. A gaze-responsive self-disclosing display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3–10. ACM, New York, NY, 1990.

J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 415–422. ACM, New York, NY, 2004.

J. Teevan, S. T. Dumais, and E. Horvitz. Potential for personalization. *ACM Transactions on Computer-Human Interaction*, 17(1):4:1–4:31, 2010.

S. Thrun. Is learning the n-th thing any easier than learning the first? In *In Advances in Neural Information Processing Systems (NIPS) 8*, pages 640–646. MIT Press, Cambridge, MA, 1996.

K. Umemoto, T. Yamamoto, S. Nakamura, and K. Tanaka. Search intent estimation from user's eye movements for supporting information seeking. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 349–356. ACM, New York, NY, 2012.

M. van Setten, S. Pokraev, and J. Koolwaaij. Context-aware recommendations in the mobile tourist application COMPASS. In *Adaptive Hypermedia and Adaptive Web-based Systems*, volume 3137, pages 515–548. Springer, Berlin, Germany, 2004.

B. M. Velichkovsky, A. Rothert, M. Kopf, S. M. Dornhöfer, and M. Joos. Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2):145–156, 2002.

S. Vemuri, C. Schmandt, and W. Bender. iRemember: a personal long-term memory prosthesis. In *Proceedings of the 3rd ACM Workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE)*, pages 65–74. ACM, New York, NY, 2006.

A. Vinokourov, J. Shawe-Taylor, and N. Cristianini. Inferring a semantic representation of text via cross-language correlation analysis. In *Advances in Neural Information Processing Systems (NIPS) 16*, pages 1473–1480. MIT Press, Cambridge, MA, 2003.

D. J. Ward, A. F. Blackwell, and D. J. C. MacKay. Dasher - a data entry interface using continuous gestures and language models. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology*, pages 129–137. ACM, New York, NY, 2000.

R. White, P. Bennett, and S. Dumais. Predicting short-term interests using activity-based search context. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pages 1009–1018. ACM, New York, NY, 2010.

R. W. White, I. Ruthven, and J. M. Jose. A study of factors affecting the utility of implicit relevance feedback. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 35–42. ACM, New York, NY, 2005.

R. W. White, J. M. Jose, and I. Ruthven. An implicit feedback approach for interactive information retrieval. *Information Processing & Management*, 42(1): 166—190, 2006.

S. Wolfram. "What are you going to do next?" Introducing the predictive interface, 2012. URL `http://blog.stephenwolfram.com/2012/12/what-are-you-going-to-do-next-introducing-the-predictive-interface/`. Retrieved Jan 30, 2013.

S. Xu, H. Jiang, and F. C. Lau. User-oriented document summarization through vision-based eye-tracking. In *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI)*, pages 7–16. ACM, New York, NY, 2009.

K. F. Yeung and Y. Yang. A proactive personalized mobile news recommendation system. In *Proceedings of Developments in E-systems Engineering (DESE)*, pages 207–212. IEEE, Los Alamitos, CA, 2010.

S. Yu, B. Krishnapuram, R. Rosales, and R. B. Rao. Bayesian co-training. *Journal of Machine Learning Research*, 12(Sep):2649–2680, 2011.

A. Zimmermann and A. Lorenz. LISTEN: a user-adaptive audio-augmented museum guide. *User Modeling and User-adapted Interaction*, 18(5):389–416, 2008.

DISSERTATIONS IN INFORMATION AND COMPUTER SCIENCE

Aalto-DD137/2012	Paukkeri, Mari-Sanna Language- and domain-independent text mining. 2012.
Aalto-DD133/2012	Ahlroth, Lauri Online Algorithms in Resource Management and Constraint Satisfaction. 2012.
Aalto-DD158/2012	Virpioja, Sami Learning Constructions of Natural Language: Statistical Models and Evaluations. 2012.
Aalto-DD20/2013	Pajarinen, Joni Planning under uncertainty for large-scale problems with applications to wireless networking. 2013.
Aalto-DD29/2013	Hakala, Risto Results on Linear Models in Cryptography. 2013.
Aalto-DD44/2013	Pylkkönen, Janne Towards Efficient and Robust Automatic Speech Recognition: Decoding Techniques and Discriminative Training. 2013.
Aalto-DD47/2013	Reyhani, Nima Studies on Kernel Learning and Independent Component Analysis. 2013.
Aalto-DD70/2013	Ylipaavalniemi, Jarkko Data-driven analysis for natural studies in funtional brain imaging. 2013.
Aalto-DD61/2013	Kandemir, Melih Learning Mental States from Biosignals. 2013.
Aalto-DD90/2013	Yu, Qi Machine Learning for Corporate Bankryptcy Prediction. 2013.

Aalto-DD 128/2013

ISBN 978-952-60-5294-6
ISBN 978-952-60-5295-3 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Information and Computer Science
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS