

Universität Duisburg-Essen
Fachbereich Informationssysteme

Eyetracking im interaktiven Information Retrieval

Seminar: Interaktives Information Retrieval

Sebastian Dungs

30. Oktober 2009

Inhaltsverzeichnis

1	Einleitung	3
1.1	Motivation	3
1.2	Funktionsweise eines Eyetrackers	3
1.3	Überblick	4
1.4	Thesen	4
2	Wahrnehmung von Retrievalsystemen	5
2.1	Betrachtung von Ergebnislisten	5
2.2	Veränderte Darstellung von Ergebnislisten	8
3	Generierung impliziten Feedbacks	9
3.1	Gewinnung impliziten Feedbacks mittels Blickbewegungsaufzeichnung . .	10
4	Relevanzanalyse auf Sub-Dokument-Ebene	12
4.1	Leseerkennungsalgorithmus	12
4.2	Bestimmung von Textrelevanz	13
5	Automatisch erstellte Annotationen	15
6	Fazit	16
	Literaturverzeichnis	17

1 Einleitung

Die vorliegende Arbeit ist im Rahmen des Seminars „Interaktives Information Retrieval“ an der Universität Duisburg-Essen im Sommersemester 2009 entstanden. Es werden die Einsatzfelder von Eyetrackern aufgezeigt. Außerdem wird darauf eingegangen, welche Vorteile sich aus deren Verwendung während des interaktiven Information Retrieval (IIR) ergeben.

1.1 Motivation

Goldstein beschreibt, dass die unterbewussten Bewegungen der Augen stark mit dem aktuellen Aufmerksamkeitsfokus korrelieren (siehe S.134 [7]). Daher können Eyetracker in realistischen Szenarien eingesetzt werden, um aufschlussreiche Informationen darüber zu gewinnen, wie die Nutzer die Systeme wahrnehmen. Durch diesen Erkenntnisgewinn eröffnen sich eine Vielzahl von Anwendungsmöglichkeiten, auf welche in dieser Arbeit detailliert eingegangen werden soll. Zunächst wird aber die grundlegende Funktionsweise eines Eyetrackers erläutert.

1.2 Funktionsweise eines Eyetrackers

Moderne Eyetracker ermöglichen das Aufzeichnen der Blickbewegungen, ohne dass die betreffende Person sich dieses Prozesses permanent bewusst ist. Während früher große Apparaturen am Kopf des Probanden befestigt oder dieser in einer bestimmten Position fixiert wurde, sind solche Maßnahme durch die fortgeschrittene Technik heute nicht mehr nötig. Deshalb kann der Anwender, während eine Aufzeichnung erfolgt, ungestört seiner Arbeit nachgehen, wodurch man sehr realistische Blickverläufe erhält.

Eyetracker arbeiten mit infrarotem Licht, wobei Lichtquellen und Kameras oftmals direkt in einen LCD-Monitor integriert sind. Sie senden das unsichtbare infrarote Licht aus und zeichnen die Reflektionen, die auf den Augen des Anwenders entstehen, mit speziellen Kameras auf. Nach aktuellem Stand der Technik müssen die Geräte für jeden Nutzer individuell konfiguriert werden. Danach ist die Software des Eyetrackers mit Hilfe eines geometrischen Modells in der Lage, aus den aufgezeichneten Reflektionsdaten die Punkte zu berechnen, die auf dem Bildschirm betrachtet wurden.

Man unterscheidet Blickbewegungen grob in Fixationen und Sakkaden. Als Fixation bezeichnet man eine Phase, in der das Auge nahezu still steht. Diese dauert typischerweise zwischen 100 und maximal 500 ms an, je nachdem welcher Beschäftigung der Mensch

nachgeht. Die Sakkaden sind „Sprünge“ zwischen einzelnen Fixationsphasen, in denen das Auge einen anderen Bereich der Umwelt, hier des Bildschirms, ansteuert. Sie dauern nur 40 - 50 ms. Entscheidend ist, dass nur während der Fixationen eine Sinneswahrnehmung stattfindet. Während dieser wird unterbewusst auch das Ziel für die nächste Fixation gesucht, welches dann durch eine Sakkade angesteuert wird.

Der Eyetracker liefert also Daten über Position und Dauer der Fixationen und zeichnet den Verlauf der Sakkaden auf. Die Blickpunkte können mit den Bildschirminhalten zusammen aufgezeichnet und anschließend ausgewertet werden.

1.3 Überblick

Im folgenden sollen kurz einige Studien und Forschungsprojekte der letzten Jahre vorgestellt werden, die Eyetracking im Kontext des interaktiven Information Retrievals einsetzen.

Viele Studien wurden zum Thema Benutzerinteraktion mit (Web-)Suchmaschinen durchgeführt. Cutrell und Guan zum Beispiel untersuchten, inwieweit sich eine veränderte Präsentation der Ergebnisliste einer Web-Suchmaschine auf das Nutzerverhalten auswirkt [5]. Von Buscher, Cutrell und Morris stammt ein Modell, das vorhersagt, welches Maß an Aufmerksamkeit ein bestimmtes Element einer Webseite erhält [1]. Auch die Frage, ob Eyetracker implizites Feedback generieren können, wurde untersucht. Hier sind die Arbeiten von Buscher, Dengel und anderen besonders interessant [12]. Ebenfalls von Buscher und Dengel stammen Forschungsarbeiten, die sich mit der Relevanzanalyse auf Sub-Dokument Ebene beschäftigen [2], oder auf automatisch aus Blickdaten generierte Annotationen eingehen [3].

Betrachtet man die genannten Arbeiten, lassen sich die folgenden interessanten Thesen herausarbeiten, auf die dann jeweils genauer eingegangen werden soll.

1.4 Thesen

- Eyetracker bieten intensive Einblicke, wie User Retrievalsysteme wahrnehmen.
- Die Verwendung von Eyetrackern im IIR erlaubt die Generierung beträchtlicher Mengen impliziten Feedbacks, welches die Retrievalergebnisse verbessern kann.
- Mittels Aufzeichnung der Blickbewegungen lassen sich auf Sub-Dokument Ebene Interessen erkennen.
- Durch Blickbewegungen automatisch erstellte Annotationen können in bestimmten Szenarien, mit vielen gleichen bzw. sehr ähnlichen Informationsbedürfnissen, hilfreich sein.

2 Wahrnehmung von Retrievalsystemen

Eyetracker dienen vor allem dazu, das Blickverhalten der Benutzer festzuhalten. Die gewonnenen Daten ermöglichen ein weitreichendes Verständnis darüber, auf welche Art und Weise die Benutzer mit dem System arbeiten. Oftmals sind Eyetracker-Studien nach dem Prinzip aufgebaut, dass ein möglichst realistisches, aber innerhalb der Studie standardisiertes, Anwendungsszenario entworfen wird. Damit sind die Ergebnisse der verschiedenen Probanden untereinander vergleichbar. Betrachtet man nun etwa zwei leicht unterschiedliche Konfigurationen des Systems, ist mit Hilfe der Eyetrackingdaten zu erkennen, mit welcher Konfiguration die Probanden besser arbeiten können. Dann kann man die Gestaltung des System eventuell entsprechend anpassen.

2.1 Betrachtung von Ergebnislisten

Im Kontext des IIR wird auch ein anderer Ansatz verfolgt. Von verschiedenen Forschungsgruppen wurden Studien mit Unterstützung durch Eyetracker durchgeführt, die sich mit Web-Suchmaschinen beschäftigten. Speziell wurde das Verhalten der Anwender untersucht, während diese die Ergebnisansicht einer Web-Suchmaschine betrachteten. Wichtig ist dabei die Darstellung der einzelnen Ergebnisse auf dem Bildschirm, nämlich als nach Relevanz geordnete Liste. Die im Folgenden beschriebenen Studien ließen sich also auch analog mit anderen Systemen durchführen, wie zum Beispiel digitalen Bibliotheken, solange die Darstellung der Ergebnisse wie oben genannt ist:

Bei mehreren Studien zeigte sich, dass das Vertrauen der Nutzer in die Relevanzbewertung der Suchmaschinen sehr hoch ist. Dies äußert sich vor allem darin, dass Anwender die Ergebnisse im oberen Bereich der Treffermenge deutlich überdurchschnittlich oft betrachten und anklicken. Dies ist selbst dann der Fall, wenn die Ergebnisse weiter unten objektiv relevanter sind.

Die Studie von Granka et al. [8] beschäftigt sich genau mit diesem Aspekt der Nutzerwahrnehmung von Ergebnislisten. Die Autoren führten eine Versuchsreihe mit 26 Probanden durch und kamen dabei zu dem Schluss, dass die Blickbewegung der Probanden beim Betrachten der Ergebnisliste generell von oben nach unten verläuft. Dies wird auf zwei Wegen veranschaulicht, wobei jeweils die zehn Ergebnisse auf der ersten Seite betrachtet werden. Zum einen wird untersucht, wie viele andere Ergebnisse sowohl oberhalb, als auch unterhalb des Ergebnisses durchschnittlich angesehen worden sind

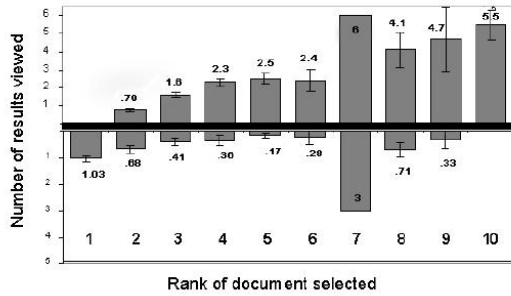


Abbildung 2.1: Anzahl der betrachteten Elemente

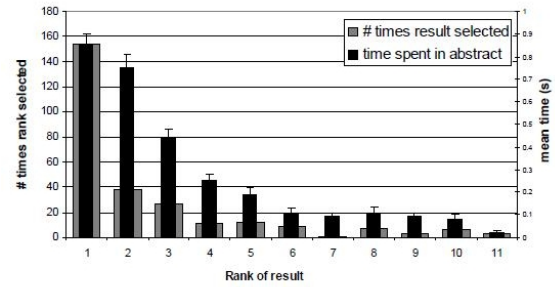


Abbildung 2.2: Betrachtungszeiten und Auswahlhäufigkeiten

und welches im Anschluss von den Probanden angeklickt wurde. Eine Übersicht dieser Auswertung ist in Abbildung 2.1 zu sehen. Es wird deutlich, dass insgesamt mehr Ergebnisse oberhalb als unterhalb der ausgewählten Links betrachtet werden. Außerdem ist der Trend zu erkennen, vor einem Klick auf ein Ergebnis weit unten in der Liste deutlich mehr andere Ergebnisse zu betrachten, als vor einem Klick auf einen Link der im oberen Bereich dargestellt wird.

Zum anderen zeigt sich deutlich, dass die Aufmerksamkeit auf der ersten Seite der Ergebnismenge bei der Websuche sehr ungleich zugunsten der ersten beiden Ergebnisse verteilt ist. Gemessen an der Betrachtungszeit erhalten diese annähernd gleich viel Aufmerksamkeit, während bereits das dritte Ergebnis weniger als zwei Drittel der Zeit betrachtet wird. Der Trend setzt sich dann weniger stark für die anderen Ergebnisse fort (Abb. 2.2). Noch deutlicher ist der Unterschied in der Auswahlhäufigkeit zwischen einzelnen Rängen. Das erste Ergebnis erhält mehr als dreieinhalb mal so viele Klicks wie das zweite. Die weiteren Ergebnisse folgen in diesem Fall mit geringerem Abstand, als bei der Betrachtungszeit. Interessant ist auch der Trend, der ab dem sechsten Rang zu erkennen ist. Die Ergebnisse sechs bis zehn werden annähernd gleich lang betrachtet und gleich oft ausgewählt. Vor dem Hintergrund, dass für gewöhnlich fünf Ergebnisse der Liste sichtbar sind, ohne dass gescrollt werden muss, erscheint diese Beobachtung durchaus plausibel.

Zwar zeigen die Auswertungen, dass die Probanden mit deutlicher Tendenz die oberen Ergebnisse bevorzugen. Dennoch geben die Autoren für keinen ihrer Befunde eine statistische Signifikanz an. Um die Erkenntnisse zu verallgemeinern, müssten daher weitere ähnliche Studien durchgeführt werden, die diesen Mangel nicht aufweisen.

Cutrell und Guan zeigten in ihrem Experiment [9], welche Auswirkungen der Rang des relevantesten Ergebnisses innerhalb der Ergebnisliste auf das Verhalten des Nutzers hat. Für dieses Experiment lässt sich festhalten, dass die Probanden signifikant mehr Zeit

Zielposition	1	2	4	5	7	8
Navigationstask	100	89	72	56	56	56
Informationstask	94	94	89	44	39	22

Tabelle 2.1: Prozentanteil der Probanden, die das Ziel betrachteten

für die Bearbeitung der Aufgaben brauchten und dabei signifikant schlechtere Ergebnisse erzielten, wenn die relevanten Ergebnisse weit unten in der Ergebnisliste angezeigt wurden. Die Autoren begründen dies unter anderem damit, dass die Benutzer während ihres Experiments kaum die Suchergebnisse betrachteten, die weit unten dargestellt waren. Dieser Zusammenhang wird in Tabelle 2.1 verdeutlicht. Mit „Zielposition“ ist dabei die Position innerhalb der Liste gemeint, an der sich das, von den Autoren für relevant befundene, Ergebnis befindet. Die Angaben verstehen sich als die Anzahl der Probanden in Prozent, die das Zielergebnis an der jeweiligen Position betrachtet haben.

Tabelle 2.1 zeigt noch eine andere Auffälligkeit. Es bestehen teilweise erhebliche Unterschiede im Benutzerverhalten in Abhängigkeit vom aktuellen Task des Benutzers. Diese Unterschiede werden von den Autoren auch anhand weiterer Zahlen beschrieben: Wenn den Probanden die Zielergebnisse weiter unten in der Ergebnisansicht präsentiert wurden – Rang 7 und 8 in diesem Experiment – sank die Erfolgsquote je nach Task unterschiedlich stark. Im Falle des Navigationstasks lösten immerhin noch 33% bzw. 22% der Probanden die Aufgabe korrekt. Bei der Informationssuche hingegen erledigten nur 6% der Probanden die Aufgabe korrekt, wenn das Ziel auf Position 7 dargestellt wurde. Lag das Ziel auf Position 8 fand kein Proband die Lösung. Die Informationssuche wird also in diesem Fall deutlich stärker durch schlechte Retrievalqualität negativ beeinflusst.

Wurde das System so konfiguriert, dass das Suchziel nicht auf den ersten beiden Positionen der Liste zu finden war, klickten die Probanden in 46% der Fälle auf das erste Ergebnis. Während einer Navigationsaufgabe schickten 15% der Probanden eine neue Anfrage ab, bei der Informationssuche jedoch nur 4%. Die geringe Anzahl an reformulierten Anfragen zeigt, besonders im letztgenannten Fall, den Anspruch der Probanden an das System. Diese gehen im Allgemeinen davon aus, dass ihr Informationsbedürfnis bereits durch die erste Suchanfrage vom System befriedigt werden kann. Dafür spricht auch hier wieder die sehr hohe Anzahl an Klicks auf dem ersten Ergebnis.

Allerdings stellten die Autoren bei ihren Untersuchungen fest, dass kein signifikanter Zusammenhang zwischen Task und Sucherfolg besteht. Da auch die Fixationszeiten bei beiden Tasks sehr ähnlich waren, liegt die Vermutung nahe, dass es sich bei den beobachteten Differenzen um zufällige Schwankungen während des Versuchs handelt. Dennoch bestätigt die Studie die oben genannte Behauptung, dass Nutzer die oberen Ergebnisse der Liste sowohl signifikant öfter betrachten, als auch bevorzugt anklicken.

Cutrell und Guan zeigten auch in anderen, verwandten Studien, dass es Unterschiede im Nutzerverhalten, je nach aktuellem Task gibt. Könnte ein System, mit Hilfe eines Eye-trackers, den Task des Benutzers erkennen, ließe sich eventuell ein Verfahren entwickeln, dass den Nutzer je nach Task optimal bei seiner Suche unterstützt.

2.2 Veränderte Darstellung von Ergebnislisten

Von den gleichen Autoren stammt auch eine Arbeit, die sich mit dem bei Websuchen üblichen kontextabhängigen Textausschnitt des jeweiligen Ergebnisses beschäftigt [5]. In einer Studie werden die Auswirkungen einer Veränderung der Länge dieses Ausschnitts untersucht, indem mit Hilfe eines Eyetrackers das Nutzerverhalten evaluiert wird.

Als Ergebnis lässt sich festhalten, dass die Länge des Textausschnitts deutlichen Einfluss auf das Blickverhalten der Nutzer hat. Je länger der Textausschnitt, desto weniger lange werden die anderen Informationen der Ergebnisseite betrachtet. Dieser Effekt führt laut den Autoren dazu, dass die Probanden beim Informationstask bessere und beim Navigationstask schlechtere Ergebnisse erzielten, wenn der Textausschnitt verlängert wurde. Als mögliche Erklärung für diese Beobachtung wird die Tatsache genannt, dass beim Navigationstask die URL des Ziels von besonderer Bedeutung ist. Diese gerät aber bei einem langen Textausschnitt zunehmend in den Hintergrund, wodurch die Auswahl von relevanten Ergebnissen erschwert wird.

3 Generierung impliziten Feedbacks

Der Begriff *Feedback* fasst die Rückmeldung des Benutzers an ein System zusammen. Gerade im interaktiven Information Retrieval ist Feedback ein wichtiger Faktor des Gesamtprozesses, denn es existieren verschiedene Ansätze, die die Rückmeldungen nutzen, um die Qualität des Retrievalsystems zu erhöhen. Im wesentlichen basieren diese Ansätze auf einer Relevanzbewertung der vom System zurückgelieferten Ergebnisse durch den Benutzer, dem sogenannten *Relevanzfeedback*. Auf eine spezielle Art von Feedback, dem *impliziten Relevanzfeedback*, soll in diesem Kapitel genauer eingegangen werden.

Grundsätzlich unterscheidet man grob zwei Arten von Feedback. Zum einen gibt es das *implizite Feedback*. Der Begriff fasst alle Reaktionen eines Benutzers zusammen, die dieser unbewusst zeigt. Beispiele für lange bekannte Maße impliziten Feedbacks sind die Nutzungszeiten oder -häufigkeiten eines Systems. Auch eine Messung der Atem- oder Herzfrequenz wurde in Experimenten dazu verwendet, Aufschlüsse über den Gefühlszustand der Probanden zu gewinnen. Mit einem Eyetracker lässt sich eine weitere Art des impliziten Feedbacks gewinnen. Die Blickbewegung und der Pupillendurchmesser geben sehr gute Hinweise auf die Aufmerksamkeitsverteilung eines Probanden [7].

Implizitem Feedback steht *explizites Feedback* gegenüber, also zum Beispiel eine Systembewertung der Anwender anhand von Fragebögen oder Umfragen. Der Vorteil von implizitem gegenüber explizitem Feedback ist nun, dass ersteres keine zusätzliche kognitive oder zeitliche Belastung für den Benutzer bedeutet. Daher ist die Menge an theoretisch verfügbarem, implizitem Feedback auch weitaus höher, beziehungsweise nur durch technische Auswertungsverfahren begrenzt. Kann man von diesem Kanal der Informationsgewinnung Gebrauch machen, lassen sich die Daten eventuell nutzen, um die betreffenden Systeme zu verbessern. Deshalb konzentrieren sich aktuelle Forschungsprojekte auch darauf, Verfahren für die Nutzung dieser Daten zu entwickeln.

Die Auswertung von implizitem Feedback ist allerdings teilweise schwierig. Die Daten unterliegen möglicherweise einem starken *Bias*. Das Klickverhalten der Benutzer von Websuchmaschinen wurde z. B. von Joachims et al. untersucht [10]. Die Autoren beschreiben einen sogenannten *Trust-Bias*, welcher durch das Vertrauen der Nutzer in das *Ranking* der Suchmaschinen zustande kommt. Vergleiche hierzu auch die Erkenntnisse aus Kapitel 2.1.

Um die Existenz und die Stärke des Trust-Bias zu untersuchen, führten Joachims et al. ein Nutzerexperiment durch. Dabei manipulierten sie die Reihenfolge der Ergebnisse auf der ersten Antwortseite einer Websuchmaschine. Der Kontrollgruppe wurde die Er-

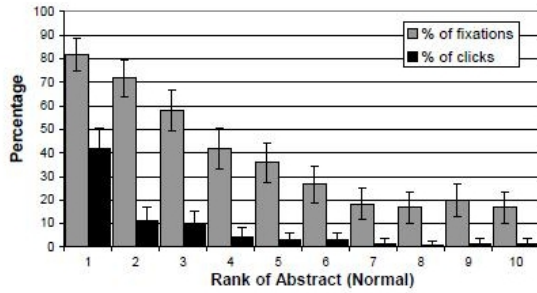


Abbildung 3.1: Ergebnisse der Kontrollgruppe

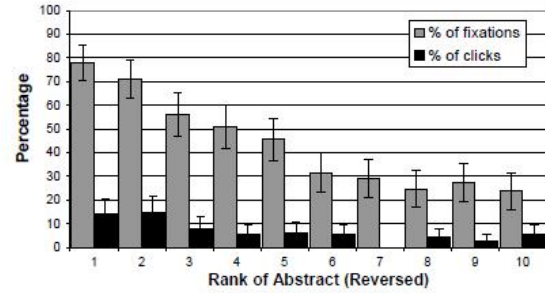


Abbildung 3.2: Ergebnisse der Experimentalgruppe

gebnisseite im Ursprungszustand dargeboten. Die Auswertung der Blick- und Klickdaten ist in Abbildung 3.1 gezeigt. Deutlich wird eine ähnliche Verteilung von Fixationen und Klicks, wie sie auch Granka et al. feststellten (Abb. 2.2).

Für die Experimentalgruppe wurde die Ergebnisliste invertiert. Die Auswertungen für diese Gruppe sind in Abbildung 3.2 dargestellt. Es existieren signifikante Unterschiede, sowohl in der Anzahl der Fixationen auf den einzelnen Ergebnissen der Liste, als auch in der Häufigkeit, mit der diese angeklickt wurden. Es zeigt sich also einerseits, dass das Verhalten der Nutzer von der Qualität des Retrievalsystems abhängig ist. Andererseits könnte man erwarten, dass die Verteilung der Klicks bei den Probanden der Experimentalgruppe annähernd umgekehrt zu der der Kontrollgruppe ist. Das dies nicht der Fall ist, kann man als Bestätigung der Theorie des Trust-Bias von Joachims et al. werten. Bei der Auswertung von Klickdaten als Relevanzfeedback ist dieser Fehler also zu berücksichtigen.

3.1 Gewinnung impliziten Feedbacks mittels Blickbewegungsaufzeichnung

Das implizite Feedback, das mit Hilfe von Eyetrackern gewonnen wird, bietet gegenüber den erwähnten anderen Arten impliziten Feedbacks den Vorteil, dass es direkt mit der Aufmerksamkeit korreliert. Wenn diese Daten von einem Retrievalsystem selbst ausgewertet werden könnten, sind sie möglicherweise auf mehrere Arten hilfreich, um das betreffende System zu verbessern.

In Kapitel 2.1 wurde auf Unterschiede im Benutzerverhalten, in Abhängigkeit vom Task eingegangen. Wenn generell Unterschiede im Verhalten auftreten, können diese von einem Retrievalsystem mit integriertem Eyetracker erkannt werden. Das System kann dann dem Task entsprechend reagieren. Zum Beispiel könnten bei einer Navigations-

che die URL auf der Ergebnisliste optisch hervorgehoben werden. Ähnliche Anpassungen sind auch in anderem Kontext denkbar.

Zukünftige Systeme könnten diesen Ansatz noch erweitern. Wenn Eyetracker eingesetzt werden, um die Arbeitsgewohnheiten von Nutzern aufzuzeichnen, kann das System für einzelne Nutzer personalisiert werden. Denkbar wäre eine dynamische Anpassung einzelner Komponenten der Systeme, basierend auf zuvor gewonnenen Blickdaten. Im Kontext des obigen Beispiels der Websuche könnte die URL oder der Textausschnitt eines Suchergebnisses je nach Präferenz des Nutzers jeweils zugunsten der anderen Elemente reduziert dargestellt oder ausgeblendet werden.

Interessant erscheint auch der Ansatz, die Blickdaten als implizites Relevanzfeedback auszuwerten. Erste Schritte in diese Richtung machten Salojärvi et al. [11]. Die Autoren beschreiben einen Eyetracker gestützten Versuchsaufbau, bei dem den Probanden zuerst eine Fragestellung auf dem Bildschirm präsentiert wurde. Als zweites Bild wurde eine Liste mit zwölf Titeln eingeblendet, aus denen die Probanden die für die Aufgabe Relevanten herausuchen sollten.

Bei der Auswertung der Daten konzentrierten sich Salojärvi et al. auf klassische Data-Mining Verfahren wie *LDA* oder *SOM*. Ziel war es, die für die Relevanzbewertung bestimmenden Blickmuster herauszuarbeiten. Entsprechend wurden zum Beispiel die Maße *mittlere Fixationsdauer* und *Regressionsdauer* als wichtig zur Unterscheidung zwischen relevanten und nicht relevanten Titeln genannt.

Die Studie unterliegt allerdings einigen Einschränkungen. So basieren alle Trainingsdaten der maschinellen Lerner auf den Blickdaten von nur drei Probanden. Die Fixationsdaten des Eyetrackers wurden manuell auf den jeweiligen Bildschirminhalt abgebildet, wobei von den Autoren keinerlei Angaben zur Genauigkeit dieser Abbildung gemacht werden. Außerdem ist das Versuchsssetup vollkommen starr, es ist keinerlei Nutzerinteraktion möglich. Man kann daher nicht von realistischen Voraussetzungen sprechen. Im nächsten Kapitel werden die Arbeiten von Buscher et al. vorgestellt, die diese Mängel nicht aufweisen.

4 Relevanzanalyse auf Sub-Dokument-Ebene

Zur Bestimmung der Relevanz eines Dokuments im Kontext des IIR, ohne Unterstützung durch Eyetracker, können Maße wie die Klickfrequenz oder die Betrachtungszeit des Dokuments verwendet werden. Diese Maße beziehen sich auf das gesamte Dokument und sind sehr ungenau. Die Betrachtungszeit hängt natürlich auch von der Dokumentgröße ab, weshalb sie nicht als absolutes Maß dienen kann. Wie oben gezeigt, ist die Bereitschaft der Nutzer auf ein Dokument zu klicken, auch von der Position desselben in der Liste der Suchergebnisse abhängig. Eine präzise Bewertung der Dokumentenrelevanz zu einer bestimmten Suchanfrage lässt sich so also nicht gewinnen. Außerdem erfolgen alle Betrachtungen anhand des gesamten Dokuments. Durch den Einsatz von Eyetrackern lassen sich genauere Analysen auf niedrigeren Ebenen, also Teilen des Dokuments wie zum Beispiel Absätzen oder Sätzen, durchführen.

4.1 Leseerkennungsalgorithmus

Distance and direction in letter spaces	Feature	Reading detector score s_r	Skimming detector score s_s
$0 < x \leq 11$	Read forward	10	5
$11 < x \leq 21$	Skim forward	5	10
$21 < x \leq 30$	Long skim jump	-5	8
$-6 \leq x < 0$	Short regression	-8	-8
$-16 \leq x < -6$	Long regression	-5	-3
$x < -16$ and y according to line spacing	Reset jump	5 and line delimiter	5 and line delimiter
All other movements	Unrelated move	Line delimiter	

Abbildung 4.1: Übersicht der *Features* des Algorithmus

Auf Grundlage der Arbeiten von Salojärvi et al. [11] und Campbell et al. [4] entwickelten Buscher et al. [12] einen sogenannten *Leseerkennungsalgorithmus* (LEA). Er stellt

einen wichtigen Schritt in Richtung automatisierter Auswertung des aus Eyetrackerdaten gewonnenen impliziten Feedbacks dar. Das Funktionsprinzip ist wie folgt:

Zuerst werden Blickpunkte nach einem genauen Verfahren zu Fixationen zusammengefasst, wenn diese die für das Lesen typische Länge von 80 ms übersteigen. Bei einer Abtastgeschwindigkeit von 50 hz sind dafür vier Blickpunkte erforderlich. Die so berechneten zusammenfassenden Fixationen werden zueinander in Beziehung gesetzt, das heißt die Übergänge zwischen den Fixationen werden in sogenannte *Features* kategorisiert. Abbildung 4.1 zeigt diese Features. In der linken Spalte ist beschrieben, wann ein Übergang einem bestimmten Feature zugeordnet wird.

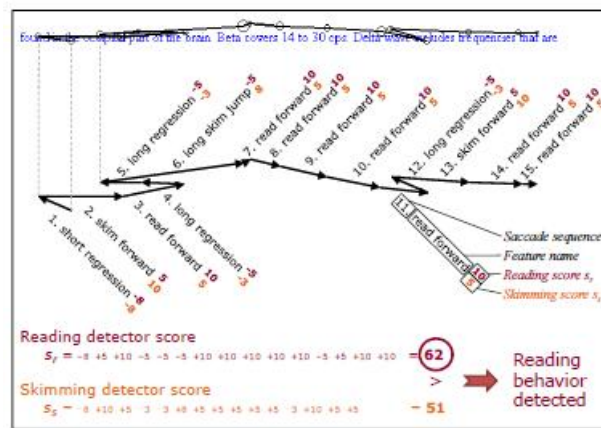


Abbildung 4.2: Blickbewegungsanalyse einer Sequenz

Im nächsten Schritt werden die Features zu Sequenzen zusammengefasst. Diese werden wiederum durch zwei spezielle „Stop“-Features beendet, siehe *Reset jump* bzw. *Unrelated move* in 4.1. Der LEA unterscheidet zwischen *lesen* und *überfliegen*. Dazu werden zwei Zähler eingeführt. Jedes Feature einer Sequenz erhöht die beiden Zähler s_r und s_s unterschiedlich stark (vgl. die rechten Spalten in Abb. 4.1). Die Sequenzen werden anhand des Zählerstand einem Blickverhalten zugeordnet. Wenn $s_s < s_r$ gilt, wird für die Sequenz *lesen* erkannt. Entsprechend wird, wenn $s_r < s_s$ gilt, die Sequenz als *überfliegen* betrachtet. Vergleiche hierzu auch Abbildung 4.2.

4.2 Bestimmung von Textrelevanz

Buscher et al. schlagen vor, die Relevanz eines Textteils bezüglich einer bestimmten Fragestellung an dem Verhältnis zwischen der Länge des gelesenen und der des überflogenen Textes zu bewerten. Sie liefern in [12] auch erste Beispiele für die Gültigkeit dieser These, wenn die individuellen Unterschiede im Leseverhalten der Benutzer bei der Auswertung berücksichtigt werden.

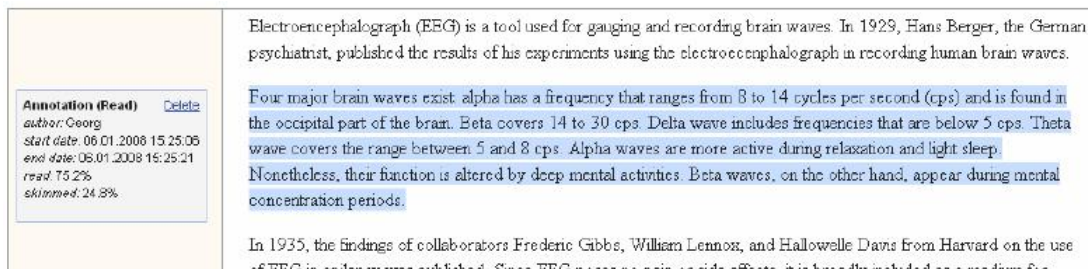
Der nächste Schritt in Richtung des eingangs erwähnten Ziels, die Retrievalqualität der Systeme zu verbessern, wird von Buscher et al. in [6] genannt. Mit einer weiteren Nutzerevaluation werden Faktoren im Blickverhalten herausgearbeitet, die für das Lesen von taskrelevanten Texten üblich sind. Die Grundlage dieses Verfahrens bildet wieder der LEA.

Bei den Ergebnissen sind Unterschiede im Blickverhalten auffällig, je nachdem, ob dem Probanden das betreffende Dokument bereits bekannt ist oder nicht. Dies führt dazu, dass bei einer Filterung der relevanten Dokumentteile mit dem LEA die Precision von 0,88 für bekannte auf 0,55 für unbekannte Dokumente fällt. Der Recall ist mit 0,84 zu 0,88 stabil. Diese Erkenntnis ist nicht weiter überraschend, wenn man bedenkt, dass beim Lesen eines unbekanntes Dokuments sicher mehr Orientierung erforderlich ist, als in bereits erschlossenen Texten. Daher ist das Blickverhalten im ersten Fall stärker verfälscht.

Die Autoren suchten daher besonders für die Fälle zusätzliche Maße, in denen unbekannte Dokumente betrachtet wurden, um relevante Textpassagen besser bestimmen zu können. Vielversprechend ist der Ansatz, zur Relevanzbestimmung nur die gelesenen Textteile einzusetzen, die länger als 200 Zeichen sind. Mit dieser Modifikation lässt sich eine Verbesserung des f-Maßes von 0,07 gegenüber der Ausgangslage erzielen. Berechnet man ein nutzerspezifisches *Wisker-Intervall* für die Regressionsanzahl, lässt sich auch damit eine Verbesserung erzielen. Wenn bei der Relevanzbestimmung nur die Textteile einbezogen werden, für die die Regressionsanzahl des Benutzers nahe seiner persönlichen unteren Intervallgrenze liegt, kann die Precision um 0,13 gesteigert werden. Dabei sinkt der Recall zwar um 0,03, dennoch ergibt sich eine Steigerung von 0,06 bezüglich des f-Maßes.

5 Automatisch erstellte Annotationen

Ebenfalls von Buscher et al. stammt die Idee, aus Blickdaten Dokumentannotationen zu generieren [3]. Zuerst werden wieder die gelesenen Textteile eines Dokuments mit Hilfe des LEA erkannt. Dabei wird z.B. für jeden Abschnitt festgehalten, wie viele Sätze gelesen oder überflogen wurden. Diese Information wird zusammen mit weiteren Metadaten, wie zum Beispiel der Zeitpunkt des Lesens oder die Verweildauer innerhalb des Dokuments gespeichert.



The image shows a document interface with a sidebar on the left and main text on the right. The sidebar contains an 'Annotation (Read)' box with a 'Delete' link and the following metadata: author: Georg, start date: 06.01.2008 15:25:06, end date: 06.01.2008 15:25:21, read: 75.2%, skinned: 24.8%. The main text is an article about EEG, with several lines highlighted in blue to indicate annotations. The highlighted text includes: 'Four major brain waves exist: alpha has a frequency that ranges from 8 to 14 cycles per second (cps) and is found in the occipital part of the brain. Beta covers 14 to 30 cps. Delta wave includes frequencies that are below 5 cps. Theta wave covers the range between 5 and 8 cps. Alpha waves are more active during relaxation and light sleep. Nonetheless, their function is altered by deep mental activities. Beta waves, on the other hand, appear during mental concentration periods.' and 'In 1935, the findings of collaborators Frederic Gibbs, William Lennox, and Hallowelle Davis from Harvard on the use of EEG in epilepsy was published. Since EEG poses no pain or side effects, it is broadly included as a medium for...

Abbildung 5.1: Ein Artikel mit Annotation

Mit einem Wiki, welches die Integration von Metadaten direkt unterstützt, wurde von den Autoren eine erste praxisnahe Implementierung von automatischen Annotationen geschaffen. Für jeden Artikel innerhalb des Wikis werden, wie oben beschrieben, Metadaten gewonnen und neben den betreffenden Textpassagen angezeigt. Ein Beispiel für eine Annotation ist in Abbildung 5.1 dargestellt.

Ein System wie von Buscher et al. beschrieben, kann in mehreren Situationen nützlich sein. Ein Benutzer kann von seinen eigenen Annotationen profitieren, wenn er zu einem späteren Zeitpunkt ein ähnliches Informationsbedürfnis hat. Wird also ein bereits gelesener Artikel im Wiki erneut aufgerufen, genügt ein schneller Scan des Dokuments, um die Stellen zu finden, die beim ersten Lesen als wichtig empfunden wurden. Anderen Nutzern könnten die Metainformationen dann angezeigt werden, wenn sie an das System gleiche oder ähnliche Suchanfragen gestellt haben, wie der Nutzer, der die Daten generierte.

6 Fazit

In dieser Arbeit wurden insgesamt vier Vorteile des Einsatzes von Eyetrackern im IIR gezeigt. Zum einen lassen sich mit Eyetrackern Verhaltensmuster der Benutzer eines Retrievalsystems erkennen, die ohne diese Geräte verborgen geblieben wären. Mehrere Studien belegen zum Beispiel, dass Nutzer einer Websuchmaschine die Ergebnisliste von oben nach unten betrachten. Daraus resultiert die oft beobachtete Bevorzugung der oberen Links, was die Anzahl der Klicks angeht. Ebenfalls zeigten sich Unterschiede im Blickverhalten, je nachdem welchen Task ein Benutzer gerade verfolgt. Diese müssten allerdings in weiteren Studien untersucht werden, um verlässlichere Aussagen zu gewinnen. Interessant sind auch die Untersuchungen, die eine Veränderung der Informationszusammensetzung von Ergebnislisten beschreiben.

Wenn Blickdaten mit Eyetrackern aufgezeichnet werden, ergibt sich ein neuer Kanal wie implizites Feedback der Benutzer gesammelt werden kann. Vorausgesetzt man kann einen Eyetracker in ein Retrievalsystem integrieren, ließe sich wahrscheinlich ein System entwickeln, das das implizite Feedback seiner Benutzer direkt verarbeitet. Damit wäre zum Beispiel eine automatische Personalisierung des Systems möglich, um sich an die Eigenarten des Benutzers anzupassen. Aber auch eine Adaption an den aktuellen Task ist denkbar, wenn sich die oben angesprochenen taskbezogenen Unterschiede verallgemeinern lassen.

Vielversprechend erscheint auch der Ansatz, mit Hilfe von Eyetrackern eine Relevanzbestimmung auf Sub-Dokumentebene vorzunehmen. Es wurde das Verfahren von Buscher et al. vorgestellt, mit dem aus Blickdaten automatisch jene selektiert werden können, die aufgezeichnet wurden, als der Nutzer einen Text gelesen hat. Außerdem wurden einige Metriken genannt, mit denen aus den Eigenschaften der Blickdaten auf die Relevanz eines Textes für eine bestimmte Aufgabe geschlossen werden kann.

Im letzten Teil wurde noch darauf eingegangen, wie aus Blickdaten mit Hilfe des Leseerkennungsalgorithmus automatisch Annotationen für Dokumente erstellt werden können. Diese können immer dann hilfreich sein, wenn viele gleiche oder sehr ähnliche Informationsbedürfnisse vorliegen.

Literaturverzeichnis

- [1] Georg Buscher, Edward Cutrell, and Meredith Ringel Morris. What do you see when you're surfing?: Using eye tracking to predict salient regions of web pages. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 21–30, New York, NY, USA, 2009. ACM.
- [2] Georg Buscher, Andreas Dengel, and Ludger van Elst. Query expansion using gaze-based feedback on the subdocument level. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 387–394, New York, NY, USA, 2008. ACM.
- [3] Georg Buscher, Andreas Dengel, Ludger van Elst, and Florian Mittag. Generating and using gaze-based document annotations. In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, pages 3045–3050, New York, NY, USA, 2008. ACM.
- [4] Christopher S. Campbell and Paul P. Maglio. A robust algorithm for reading detection. In *PUI '01: Proceedings of the 2001 workshop on Perceptive user interfaces*, pages 1–7, New York, NY, USA, 2001. ACM.
- [5] Edward Cutrell and Zhiwei Guan. What are you looking for?: An eye-tracking study of information usage in web search. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 407–416, New York, NY, USA, 2007. ACM Press.
- [6] Georg Buscher; Andreas Dengel. Gaze-based filtering of relevant document segments. In *Workshop on Web Search Result Summarization and Presentation. Workshop on Web Search Result Summarization and Presentation (WSSP-2009), located at in conjunction with WWW '09 18th International World Wide Web Conference, April 20-24, Madrid, Spain. o.A., 2009.*
- [7] E. Bruce Goldstein. Wahrnehmungspsychologie – Der Grundkurs. Spektrum, 2008.
- [8] Laura A. Granka, Thorsten Joachims, and Geri Gay. Eye-tracking analysis of user behavior in WWW search. In *SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 478–479, New York, NY, USA, 2004. ACM.
- [9] Zhiwei Guan and Edward Cutrell. An eye tracking study of the effect of target rank on web search. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 417–420, New York, NY, USA, 2007. ACM.

- [10] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, and Geri Gay. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Trans. Inf. Syst.*, 25(2), April 2007.
- [11] Jarkko Salojärvi, Ilpo Kojo, Jaana Simola, and Samuel Kaski. Can relevance be inferred from eye movements in information retrieval? 2003.
- [12] Georg Buscher; Andreas Dengel; Ludger van Elst. Eye movements as implicit relevance feedback. In *Proceedings and Extended Abstracts of the 26th Conference on Human Factors in Computing Systems. 26th Conference on Human Factors in Computing Systems (CHI-2008), in Conjunction with Works in Progress, April 5-10, Florence, Italy*, pages 2991–2996. ACM, 2008.