



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

DEPARTMENT INFORMATION

Forschungsarbeit

Query Understanding in EconBiz.

Wie suchen die Nutzer von EconBiz? Eine Analyse der Suchlogs des Suchportals der Zentralbibliothek für Wirtschaftswissenschaften

vorgelegt von

Alexandra Svantje Linhart

Studiengang Information, Medien, Bibliothek

erster Prüfer: Prof. Dr. Dirk Lewandowski
zweiter Prüfer: Dr. Timo Borst

Hamburg, März 2015

Inhaltsverzeichnis

Abbildungsverzeichnis	v
Tabellenverzeichnis	vi
Einleitung	1
1 Forschungsfragen	3
1.1 Suchtaktiken	3
1.1.1 Wie ändern sich die Suchtaktiken innerhalb von Sessions?	3
1.2 Known-Item-Suche	4
1.2.1 Welchen Anteil haben Known-Item-Suchen an den Suchen in EconBiz?	4
1.2.2 Mit welchen Elementen suchen Nutzer bei ihrer Known-Item-Suche?	4
1.2.3 Welche Suchtaktiken verfolgen Nutzer bei Known-Item-Suchen?	4
1.2.4 Wie erfolgreich sind die Nutzer bei ihrer Known-Item-Suche?	4
1.3 Änderungen des Sucherverhaltens in Abhängigkeit vom Einstiegspunkt	4
1.4 Dauer und Länge	5
2 Über EconBiz	6
2.1 Suchfunktionen von EconBiz	6
3 Bisherige Untersuchungen	8
3.1 Nutzerbefragungen	8
3.2 Usability-Tests	10
3.3 Logfile-Analysen	10
3.4 EconDesk - Voruntersuchung	11
4 Methodik	12
4.1 Möglichkeiten und Grenzen der Logfile-Analyse	12
4.2 Datengrundlage	13
4.2.1 Der Logfile von EconBiz	13
4.2.2 Zuordnung von Sessions	14
4.2.3 Datenreinigung	14
4.2.4 Gründe für die Aussortierung von Daten	15

4.3	Stichprobenbildung	16
4.4	Datenerfassung	17
4.4.1	Objekte	17
4.4.2	Aktionen	18
4.4.3	Erfolgsbewertung	19
5	Allgemeine Ergebnisse	22
5.1	Suchanfragen in EconBiz	22
5.1.1	Häufigkeit des Vorkommens	22
5.1.2	Länge der Suchanfragen	24
5.2	Sessiondauer	24
5.2.1	Verweildauer auf Suchergebnisseiten	25
5.3	Bounce-Sessions: Sessions mit nur einem Seitenaufruf	25
5.4	Rebounding	26
5.5	Sichtbare Klicks und Trefferzahlen	27
5.6	Weiterblättern	28
5.7	Ajax-Aktionen	28
6	Suchtaktiken	30
6.1	Definitionen von Suchtaktiken	30
6.1.1	Einstiegstaktiken	30
6.1.2	Reformulierungstaktiken	31
6.1.3	Suchfunktionstaktiken	34
6.2	Bisheriger Forschungsstand zu Suchtaktiken	36
6.3	Ergebnisse zu Suchtaktiken	37
6.3.1	Reformulierungstaktik	37
6.3.1.1	Verwendung unterschiedlicher Reformulierungstaktiken	37
6.3.1.2	Reduzieren, Erweitern oder Wechseln?	38
6.3.1.3	Sucherfolg der Reformulierungstaktiken bei thematischen Suchen	39
6.3.2	Chaining-Taktiken	40
6.3.2.1	Sucherfolg bei Chaining-Taktiken	40
6.4	Chaining vs. Suchschlitzsuche	41
6.5	Transaktionsorientiertes Chaining?	42
6.6	Filternutzung	44
6.6.1	Stichprobe	44

6.6.2	Annäherung über die Gesamtdaten	45
6.7	Fazit zu Suchtaktiken	46
7	Known-Item-Suche	48
7.1	Erkennen einer Known-Item-Suche	48
7.2	Aktueller Forschungsstand zu Known-Item-Suchen	50
7.3	Ergebnisse	52
7.3.1	Anteil der Known-Item Suchen	52
7.3.1.1	In Suchmaschinen indexierte Suchergebnisseiten	52
7.3.2	Elemente der Suchanfrage bei Known-Item-Suchen	53
7.3.2.1	Alleinstehenden Elemente einer Known-Item-Suchanfrage	55
7.3.2.2	Mehrere Elemente in Known-Item-Suchanfragen	55
7.3.2.3	Elementkombinationen in Known-Item-Suchanfragen	56
7.3.2.4	Known-Item-Suchen ohne Titel, Autor und Jahresangaben	56
7.3.3	Suchtaktiken bei Known-Item-Suchen	57
7.4	Sucherfolg bei Known-Item-Suchen	58
7.4.1	Ranking	59
7.4.2	Gründe für den Misserfolg	61
7.4.2.1	Unpräzise Suchanfragen	61
7.4.2.2	Fehlende Indexierung	61
7.4.2.3	Rechtschreibfehler	62
7.4.2.4	Falsche Feldbeschränkung	62
7.4.2.5	Sucherfolg mit Suchanfrage-Elementen	63
7.4.3	Suche nach bereits gefundenen Treffern	63
7.5	Fazit zu Known-Item-Suchen	64
8	Herkunft der Nutzer und Einstiegspunkte in EconBiz	66
8.1	Hypothesen und Forschungsstand zu Einstiegspunkten	66
8.2	Einstiegspunkte	67
8.3	Herkunft	67
8.3.1	Ankunft über Suchmaschinen	67
8.3.2	Ankunft über zbw.eu	68
8.3.3	Ankunft über Multiplikatoren	68
8.4	Absicht oder „Hineinstolpern“?	68
8.4.1	Erste Aktionen nach Hineinstolpern	69
8.4.1.1	Ankunft auf einem Einzeltreffer über eine Suchmaschine	69

8.4.1.2	Ankunft auf einer Suchergebnisseite über eine Suchmaschine	70
8.5	Verhalten nach Einstiegspunkten	70
8.5.1	Sessionlänge	70
8.5.2	Bouncing	71
8.5.3	Weiterblättern	72
8.5.4	Nutzung des Suchschlitzes	72
8.5.5	Suchtypen	73
8.5.6	Reformulierungstaktiken	74
8.5.7	Chaining-Taktiken	74
8.5.8	Filternutzung	75
8.5.9	Sucherfolg	75
8.6	Fazit zu den Einstiegspunkten	76
9	Fazit	78
9.1	Zusammenfassung	78
9.2	Grenzen der Untersuchung	80
9.3	Weitere Untersuchungen	81
	Literaturverzeichnis	83

Abbildungsverzeichnis

5.1	Absolute Häufigkeiten der Länge von Suchanfragen in Worten	24
6.1	Häufigkeiten der Reformulierungstaktiken	38
6.2	Anteile der groben Reformulierungstaktiken	38
6.3	Erfolgs- und Misserfolgsquoten bei Reformulierungen	40
6.4	Häufigkeiten der Chaining-Taktiken	41
6.5	Entstehung von Suchergebnisseiten	42
6.6	Häufigkeit von Filtertypen	44
7.1	Suchanfragen auf in Google indexierten Suchergebnisseiten	53
7.2	Häufigkeiten der Elemente in Known-Item-Suchanfragen	54
7.3	Häufigkeit von Elementkombinationen	55
7.4	Anteile von Reformulierungen bei Elementen von Known-Item-Suchanfragen .	58
7.5	Aktionen bei Known-Item-Suchen	59
7.6	Success@n für Known-Item-Suchanfragen	60
7.7	Anteil der Elemente in Known-Item-Suchanfragen nach Erfolg	63
7.8	Rel. Anteile der Misserfolgfälle und Indexierungszustand in Transaktionen, pro Element	64

Tabellenverzeichnis

4.1	Gründe für die Aussortierung beim Bot Filtern	16
5.1	Verwendung der Aktionsleisten: Ajax-Aktionen	28
6.1	Filtertypen beim 1. Schritt der Filterung	45
8.1	Längen und Dauern der Sessions nach Einstiegspunkt	71
8.2	Bouncing-Verhalten nach Einstiegspunkten	71
8.3	Weiterblättern-Verhalten nach Einstiegspunkten	72
8.4	Nutzung des Suchschlitzes nach Einstiegspunkten	73
8.5	Suchtypen nach Einstiegspunkten	73
8.6	Reformulierungsverhalten nach Einstiegspunkten	74
8.7	Chaining-Verhalten nach Einstiegspunkt	74
8.8	Filterverhalten nach Einstiegspunkten	75
8.9	Sucherfolg nach Einstiegspunkten	76

Einleitung

Suchmaschinen definieren die Standards, wie heutzutage nach Informationen gesucht wird (Nielsen (2005)). Nicht nur der Alltagsnutzer, auch in der Wissenschaft wird das Verhalten von Suchmaschinen geprägt. Die Zufriedenheit der Nutzer mit einem Suchsystem ist stark abhängig von ihren Erwartungen an das System. Um ein Suchsystem besser auf seine Nutzer einzustellen, ist es sinnvoll zu untersuchen, wie die Nutzer mit dem bisherigen System umgehen und wie erfolgreich sie dabei sind. So erfährt man auch etwas über die Stellen, an denen sie Nutzer mit ihrem Verhalten nicht mehr erfolgreich sind, oder wo sie auf Probleme bei der Nutzung stoßen.

Das Sucherverhalten der Nutzer von EconBiz, dem Suchportal der Zentralbibliothek für Wirtschaftswissenschaften, wurde in dieser Arbeit in den Blick genommen. Das Ziel der Untersuchung war ein besseres Bild vom Suchverhalten der Nutzer zu erhalten. Insbesondere ihre Suchtaktiken und der Sucherfolg oder Nicht-Erfolg wurden betrachtet. Hierzu wurden die Logfiles aus einem Jahr EconBiz Nutzung ausgewertet. Ein Logfile bietet ein Abbild von „realen“ Interaktionen im normalen Nutzungskontext, ohne dass Eingriffe in das Verhalten durch die Beobachtung nötig sind. Die Logfile-Analyse ist eine Methode, die es erlaubt eine große Menge von Daten über das Interaktionsverhalten der Nutzer zu erhalten, und so ein gutes Bild davon zu bekommen, was Nutzer häufig und was sie weniger häufig machen und wie sie unter welchen Umständen reagiert haben.

Neben einer Suche, die darauf abzielt ein besseres Bild über ein bestimmtes Thema zu erhalten, sind Bibliothekskataloge ursprünglich dafür ausgelegt worden, ein bereits bekanntes Medium in der Sammlung zu lokalisieren. Auch in der Welt der Suchmaschinen gibt es diese Art der Suche, nach Broder „navigationsorientierte“ Suchanfragen, d.h. Anfragen, bei denen das Ziel des Nutzer ist, sich auf ein bestimmte Webseite zu begeben (Broder 2002, 5). Übertragen auf den Kontext eines Online-Suchsystems für wissenschaftliche Literatur gilt dies für Suchen, bei denen der Nutzer bereits weiß, welches Dokument er finden möchte, die sogenannte „Known-Item-Suche“.

Auf eine in der Suchmaschinenforschung oft angewendete Auftrennung der Suchintentionen nach navigations-, transaktions- und informationsorientierten Anfragen wurde hier verzichtet, im Wesentlichen aus dem Grunde, dass eine saubere Unterscheidung zwischen navigations- und transaktionsorientierten Anfragen im Kontext eines Logfiles kaum möglich ist. Außerdem wäre die Definition einer „Transaktion“ im Kontext von EconBiz das Herunterladen eines

Textes oder die Ausleihe eines Buches, was höchstwahrscheinlich im Grunde das Ziel der meisten Suchen in EconBiz sein dürfte. Es wurde also nur zwischen zwei Typen unterschieden: informationsorientierte (thematische) Suche und navigationsorientierte Suche (Known-Item-Suche).

Im Folgenden werden die einzelnen Forschungsfragen und Hypothesen dargelegt. Es werden eine Reihe von Untersuchungen vorgestellt, die sich bereits mit dem Suchverhalten der Zielgruppe der ZBW oder anderen Informationssystemen innerhalb der ZBW beschäftigt haben. Daran anschließend folgt eine Darstellung der Vor- und Nachteile der verwendeten Methode der Logfile-Analyse sowie eine Übersicht über das Vorgehen in dieser Untersuchung, ehe die Ergebnisse der Untersuchung aufgeführt werden.

Kapitel 1

Forschungsfragen

Das Ziel dieser Untersuchung war es, ein Bild vom Suchverhalten der Nutzer von EconBiz zu erhalten. Die Analyse erfolgte auf Basis von Sessions, nicht auf Basis einzelner Anfragen. Hierzu wurden einzelne Server-Anfragen anhand bestimmter Kriterien zu Sessions zusammengefasst (s. 4.2.2). Es wurde versucht, ein möglichst umfassendes Bild des typischen Suchverhaltens in EconBiz zu erreichen, um Rückschlüsse auf eventuelle Problemstellen bei der Nutzung zu erhalten. Aus diesem Grund wurden in dieser Untersuchung mehrere Aspekte des Suchverhaltens in den Blick genommen:

1. Suchtaktiken
2. Known-Item-Suchen
3. Änderungen des Suchverhaltens in Abhängigkeit vom Einstiegspunkt

1.1 Suchtaktiken

Welche Suchtaktiken werden von den Nutzern verwendet?

Zu Suchtaktiken können unter anderem gezählt werden, Erweiterungen oder Verengungen der Suchbegriffe, Neu- oder Umformulierungen der Anfrage, der Umstieg auf andere Suchfelder. In EconBiz müssen diese Taktiken nicht immer zwangsläufig über die Suchanfrage direkt gelöst werden, auch die Nutzung der Filterungsmöglichkeiten ist eine Möglichkeit, eine Verengung der Suchanfrage vorzunehmen. Mit Taktiken wurde ergebnisoffen umgegangen. Es wurde also nicht im Vorhinein festgelegt, welche Taktiken vorkommen konnten.

1.1.1 Wie ändern sich die Suchtaktiken innerhalb von Sessions?

Um zu einem befriedigenden Ergebnis zu kommen, kann es auch möglich sein, dass innerhalb einer Session die Nutzer zu mehreren unterschiedlichen Taktiken greifen. Es wurde vermutet, dass sich hieran erkennen lässt, wo das System den Nutzer im Verlauf des Suchprozesses unterstützen besser könnte und wo der Nutzer auf Probleme stößt.

1.2 Known-Item-Suche

1.2.1 Welchen Anteil haben Known-Item-Suchen an den Suchen in EconBiz?

Diese Frage sollte den Stellenwert dieser Art der Suche für die Nutzung von EconBiz klären. Ein hoher Anteil an Known-Item-Suchen würde auch die Priorität für eine Verbesserung der Anfrageinterpretation von EconBiz in diesem Bereich bedeuten.

1.2.2 Mit welchen Elementen suchen Nutzer bei ihrer Known-Item-Suche?

Um besser zu verstehen, wie Nutzer bei einer Known-Item-Suche vorgehen, müssen auch die Suchanfragen genauer betrachtet werden. Auf der Ebene der Suchanfragen wurde daher untersucht, welche Elemente die Suchanfragen einer Known-Item-Suche enthalten, z.B. Jahresangaben, Autorennamen etc. Das Auftreten bestimmter Angaben könnte eventuell hilfreich sein, um diesen Typus der Suche automatisch zu identifizieren und das Ranking daran auszurichten.

1.2.3 Welche Suchtaktiken verfolgen Nutzer bei Known-Item-Suchen?

Für den Fall, dass Nutzer nicht gleich mit der ersten Known-Item-Suchanfrage innerhalb einer Session erfolgreich sind, lohnt es sich zu schauen, welche Taktiken sie einsetzten bzw. auch, welche davon Erfolg und welche keinen Erfolg haben.

1.2.4 Wie erfolgreich sind die Nutzer bei ihrer Known-Item-Suche?

Der Erfolg und die Zufriedenheit der Nutzer mit EconBiz im Falle einer Known-Item-Suche hängt maßgeblich davon ab, wie gut EconBiz in diesen Fällen darin ist, das gesuchte Item hoch zu ranken. Es wurde also versucht zu überprüfen, wie gut EconBiz im Falle der Known-Item-Suchen funktionierte. Grenzen lagen hier vor allem in den fehlenden Daten zum Ranking zum Zeitpunkt der Sessions. Es konnten nur nachträgliche Trefferzahlen und Rankings ermittelt werden.

1.3 Änderungen des Sucherverhaltens in Abhängigkeit vom Einstiegspunkt

Es wurde zunächst ermittelt, auf welche Weise die Nutzer bei ihrer Recherche in EconBiz einsteigen. Der Weg des Nutzers hin zu dem System kann Aufschluss darüber geben, in

welchem Kontext er Suchanfragen an das System stellt. Kommt er über eine andere Suchmaschine, fehlt ihm evtl. das Wissen um das genaue Angebot von EconBiz (seine Inhalte und sein Leistungsumfang). Darüber hinaus steigt er vielleicht auf einer Ebene in das Suchsystem ein, die nicht unbedingt intendiert war, was zu fehlendem Kontext führen könnte. Kommt es eventuell zu mehr Problemen in der Interaktion wenn der Sucheinstieg nicht auf der Startseite von EconBiz liegt? Es wurde untersucht, wie sich das Verhalten der Suchenden in Abhängigkeit von den Einstiegspunkten der Suche veränderte.

1.4 Dauer und Länge

Ursprünglich war eine weitere Forschungsfrage zur Dauer der Suchen angedacht: Wie lange brauchen die Nutzer, um ihr Informationsbedürfnis zu befriedigen? Der Gedanke war, dass Nutzer, die länger auf Ergebnisseiten verweilen und mehr Suchergebnisse ansehen, oder öfter die Anfragen reformulieren, Probleme haben ihr Informationsbedürfnis zu befriedigen (Aula u. a. 2010, 41). Dafür sollten Interaktionen auf Suchergebnisseiten und die zeitliche Länge der Sessions, nebst der Anzahl der Suchanfragen zu einem Thema beobachtet werden. Im Rahmen der Untersuchung ergaben sich jedoch eine Reihe von Problemen. So ist nur ein Bruchteil der Interaktionen auf den Suchergebnisseiten im Logfile erkennbar (s. 4.4.3), es können hier also keine verlässlichen Daten erhoben werden. Zweitens war die Wahl der Datenverarbeitung als relationale Datenbank sehr ungünstig für einen Vergleich von Verweildauern. Hierfür hätten Zeilenvergleiche gemacht werden müssen, was die Datenbank und den Server teilweise stark überforderte und zu Abbrüchen führte. Drittens ist es innerhalb einer Logfile-Analyse eigentlich unmöglich festzustellen, wann und ob eine Befriedigung beim Nutzer auftritt, insbesondere im Fall von thematischen Suchen. Aus diesen Gründen wurde diese Forschungsfrage im Rahmen dieser Untersuchung fallen gelassen.

Die genauen Hypothesen und den Stand der Forschung zu den einzelnen Bereich wird jeweils im Vorfeld zu den Ergebnissen der jeweiligen Fragestellungen dargelegt. Im Folgenden wird zunächst ein Überblick über EconBiz und seine Funktionen gegeben.

Kapitel 2

Über EconBiz

EconBiz ist das Fachportal der Zentralbibliothek für Wirtschaftswissenschaften. Es bietet eine Suche über die indextierten Publikationen aus dem wirtschaftswissenschaftlichen Bereich aus unterschiedlichen Datenquellen. Der Katalog der ZBW ist hier durchsuchbar, sowie die Publikationen aus dem Open-Access Repository EconStor, aus RePEc, BASE, und der Bestand der Universitäts- und Stadtbibliothek Köln, die vor der ZBW das Sondersammelgebiet Wirtschaftswissenschaften hatte (ZBW (2014)).

Ursprünglich war EconBiz die virtuelle Fachbibliothek, die von der ZBW aufgebaut wurde. Sie wurde zu einer Suchmaschine weiterentwickelt und läuft mit der Suchtechnologie vuFind (Lucene/SolR). EconBiz enthält zum Zeitpunkt der Untersuchung über 9 Mio. indextierte Publikationen.

Neben der Homepage und einigen Hilfeseiten, gibt es in EconBiz die Suchergebnisseiten, und Seiten für die einzelnen Treffer, mit Metadaten zu den indextierten Publikationen.

Die Inhalte von EconBiz sind auch in anderen Suchmaschinen anderen Suchmaschinen indextiert. Nicht nur die einzelnen Publikationen, sondern ganze Suchergebnisseiten werden von ihnen indextiert. Neben dem Suchschlitz auf der Startseite von EconBiz, ist die Suche auch in die Webseite der ZBW integriert, sowie in den Webseiten einiger Partner der ZBW.

2.1 Suchfunktionen von EconBiz

Neben der Eingabe von Suchanfragen ist es möglich, das Feld, in dem die Anfrage gesucht werden soll, zu beschränken. Auch eine erweiterte Suchfunktion wird angeboten. Auf den Suchergebnisseiten bietet EconBiz dem Nutzer eine Reihe von Facettenfiltern an, mit deren Hilfe die Ergebnismenge gezielt eingegrenzt werden kann. Des Weiteren ist es möglich eine Suchanfrage von vornherein nur auf frei verfügbare Volltexte zu begrenzen.

Sofern ein Volltext frei verfügbar ist, kann er direkt von der Suchergebnisseite aufgerufen werden. Daneben bietet EconBiz die Funktion, eine Suche nach dem Treffer in anderen Bibliotheken auszuführen. Diese Suche erfolgt dann über das Hochschulbibliothekszentrum Nordrhein-Westfalen.

Neben der Möglichkeit, die Ergebnisse umzusortieren (etwa nach Datum), können die Nutzer einzelne Treffer auch in einer Merkliste abspeichern. Hierzu ist ein Login nötig.

Sowohl auf den Suchergebnisseiten, als auch auf den Einzeltrefferseiten ist es möglich durch den Klick auf eine bibliographische Angabe (etwa den Titel der übergeordneten Publikation eines Zeitschriftenartikels) eine Suche auszulösen. EconBiz sucht dann nach allen passenden Treffern und bietet dem Nutzer so die Möglichkeit, weitere interessante Ergebnisse zu seinem Thema zu finden. Hierfür wird automatisch eine Suchanfrage aus dem entsprechenden Metadatum gebildet und die Feldbeschränkung der Suche auf das entsprechende Feld begrenzt. Auch eine Suche aus dem Standard Thesaurus Wirtschaft heraus ist möglich.

Auf den Einzeltrefferseiten bietet EconBiz zudem zusätzlich die Funktionalität an, die bibliographischen Angaben in verschiedene Zitierformate zu exportieren. Auch gibt es hier Vorschläge für „ähnliche Treffer“. Hier kann entweder zwischen verschiedenen Treffer gewechselt werden, oder eine Suchergebnisseite mit allen „ähnlichen Treffern“ aufgerufen werden.

Neben der Suche nach Publikationen gibt es eine separate Suchmöglichkeit für Veranstaltungen im Bereich der BWL/VWL, die in dieser Untersuchung nicht betrachtet wurde.

Seit Januar 2015 gibt es eine ZBW-Ansicht, bei der man die Signaturen und den Ausleihstatus aus dem OPAC der ZBW dazu schalten kann. Diese Funktion stand den Nutzer aus dem untersuchten Zeitraum jedoch nicht zur Verfügung.

Der folgende Abschnitt beleuchtet die bisherigen Untersuchungen und Kenntnisse zur Nutzergruppe der ZBW im Allgemeinen und zur Nutzung der Angebote der ZBW.

Kapitel 3

Bisherige Untersuchungen

In einer Reihe von Untersuchungen hat die ZBW sich mit ihrer Nutzergruppe beschäftigt. Im Folgenden werden die Erkenntnisse aus den verschiedenen Studien zusammengetragen. Der Fokus der einzelnen Untersuchungen war immer etwas anders gesetzt, dennoch lässt sich etwas über die Nutzer im Allgemeinen lernen.

3.1 Nutzerbefragungen

Im Jahr 2010 ließ die ZBW eine Untersuchung durchführen, mit dem Ziel die Bedürfnisse der Nutzergruppen im Bereich des Informationsmanagements zu erfassen. Zwei Jahre zuvor wurden die Dienstleistungen der ZBW in einer Befragung evaluiert. Im Jahre 2012 wurden diese Analysen dann nochmals zusammengefasst.

In der Nutzerstudie von 2010 wurden die Zielgruppen „Forschende“, „Studierende“ und „Unternehmensvertreter“ gebildet. Sie wurden in drei Stufen untersucht, zunächst mit Interviews, dann in einer Art Online-Gruppendiskussion und abschließend mit einer Online-Befragung, die insgesamt 244 Teilnehmer hatte. Gefragt wurde neben den Bedürfnissen auch nach den üblichen Recherchetaktiken, Sucheinstiegen und Problemen, mit denen die Nutzer sich bei der Recherche konfrontiert sahen. Daneben wurde nach den Hauptinteressen bei den Angeboten der ZBW gefragt, und weitere Vorschläge für neue Dienstleistungen vorgebracht, um diese von den Nutzern bewerten zu lassen.

Im Zusammenhang mit der Untersuchung der Logfiles von EconBiz ist vor allem interessant, was die Nutzer gerne suchen, wo sie Probleme bei der Suche entwickeln und welche Suchtaktiken sie verfolgen. Die Studie ergab, dass die meisten Nutzer ihren Einstieg über Google wählten, insbesondere wenn sie nach Fakteninformationen recherchierten. Bei Forschern und Studierenden kann dies für Literaturrecherchen auch Google Scholar sein, oder einschlägige Literaturdatenbanken wie EBSCO und JSTOR. Die Unternehmensvertreter hingegen nutzen Google Scholar nicht.

Gründe für die Nutzung von Google wurden angegeben mit dem guten Relevanzranking von Google und der Fähigkeit bei unvollständigen oder fehlerhaften Suchanfragen noch relevante

Ergebnisse zu liefern. Kritisiert wurde dagegen, dass der Aufwand der Suche bei Google stark schwankt zwischen sehr schnellen und langen, ergebnislosen Suchen. An Google Scholar gefiel besonders die Fähigkeit, auch nicht über die üblichen Wege veröffentlichte Arbeiten zu finden, sowie die Unterstützung bei der Relevanzbewertung durch die Angabe der Zitationszahlen. Hier hingegen wurde wiederum kritisiert, dass gute Treffer auch mal weit hinten landen oder der Zugriff auf gefundene Dokumente nicht immer gegeben ist.

Den Nutzer wurde damals auch eine Beta-Version von EconBiz vorgelegt, woran ihnen besonders die Möglichkeit der Ergebniseingrenzung durch Filter gefiel. Weniger gut wurde damals die (Tipp-)Fehlertoleranz und die Kombinierbarkeit von Suchworten bewertet.

Was die Nutzer recherchierten, hing stark von der Zielgruppe ab. So suchten Unternehmensvertreter im Wesentlichen nach Fakteninformationen oder Daten, während Forschende die neuesten Fachartikeln oder Working Papers recherchierten. Sie gaben auch an, dass sie generell Probleme haben, die Relevanz eines Treffers zu bewerten. Daneben fiel es ihnen schwer, die richtigen Suchwörter zu finden, um die besten Treffer zu erhalten.

In der Untersuchung zeigte sich auch, dass die Nutzer sich kaum Gedanken darüber machen, wie die Suche an sich funktioniert und teilweise davon ausgingen, es gäbe nichts zu ihrem Thema, wenn sie nicht in der Lage waren, etwas zu finden (Stegemann u. Swanson (2010)). In der Evaluierung der Serviceleistungen der ZBW von 2008 wurden die untersuchten Nutzer wiederum in drei Gruppen geteilt: Wissenschaftler, Studierende und Fachreferenten. Die Fachreferenten wurde als Multiplikatoren für die Dienste der ZBW in die Evaluierung mit eingeschlossen. Methodisch wurden 30 Interviews mit Vertreter dieser Gruppen geführt.

Hier zeigte sich, dass es die Nutzer wenig interessiert aus welcher Datensammlung ihre Ergebnisse stammen, und von der Fülle möglicher Zugriffspunkte überfordert waren. Meist waren ihnen die Namen der Datenbanken, die sie für Recherchen verwendeten gar nicht bekannt. Sie nutzten Google als Einstieg und kannten dann oft nur die Datenbanken, die sie dort aufgefunden hatten. Daher unterstellten sie Google eine hohe Abdeckungsrate an Fachinformation. (Genicus (2008)).

Eine kurze Online-Befragung speziell der Nutzer von EconBiz wurde im März 2014 ausgeführt. Es wurde nach dem Fachhintergrund und dem Beruf gefragt. Etwa 13.500 Mal wurden die Fragen, die als Pop-up bei der Nutzung von EconBiz gestellt wurden, beantwortet. 69 % der Antwortenden waren Studenten und weitere 11,6 % Doktoranden oder wissenschaftliche Mitarbeiter. Ein Großteil der Antwortenden war von einer nicht-deutschen IP-Adresse aus in EconBiz befragt worden (57 %). Nutzer mit ökonomischer Fachrichtung machten zusammen 62 % der Befragten aus. Ein gutes Viertel hatte keinen ökonomischen Hintergrund und 11,4 % machten keine Angaben (Pianos (2014)).

3.2 Usability-Tests

Neben diesen Nutzerbefragungen wurden einige Usability-Tests an digitalen Angeboten der ZBW gemacht. Im Jahr 2012 wurde ein Labeling-Test für EconBiz durchgeführt. Durch diesen Test ließen sich einige mögliche Missverständnisse bei der Benennung verschiedener Funktionen aufdecken. Die Labels wurden sowohl auf der Homepage, den Suchergebnisseiten als auch auf den Einzeltrefferseiten getestet. Gerade auf den Einzeltrefferseiten wurde bei den Labels jedoch nicht untersucht, welche Funktionen die Nutzer hinter einem Klick auf die entsprechenden Metadatenfelder vermuteten, sondern lediglich, ob das Label verständlich machte, um was für ein Metadatum es sich handelte (Linek u. a. (2012)). Aufgrund dieses Test wurden einige Benennungen geändert.

Ein weiterer Labeling-Test auf der Homepage und den Suchergebnisseiten von EconBiz wurde von einer HAW-Studentin 2013 durchgeführt. Auch hier kam man zu Verbesserungsvorschlägen für die Veränderung der Labels in EconBiz. Leider wurden hier nicht die Labels auf Einzeltrefferseiten getestet (Friedrichsen (2013)).

3.3 Logfile-Analysen

Neben den Untersuchungen durch Befragungen und Usability-Tests wurden bereits ein paar Logfile-Analysen von Econis, dem Katalog und EconStor, dem Repository durchgeführt. Das Hauptaugenmerk bei diesen Analysen richtete sich auf die mögliche Abdeckung der Suchanfragen der Nutzer mit Begriffen aus dem Standard Thesaurus Wirtschaft, oder anderen kontrollierten wirtschaftswissenschaftlichen Vokabularen. Die Analyse der Econis-Logs wertete 1,5 Millionen Suchanfragen aus 1,5 Jahren Econis Logs aus. Sie ergab eine breite Streuung der eingegebenen Suchstrings, von denen mehr als die Hälfte nur ein einziges Mal in dieser Form vorkamen. Am häufigsten wurden zwei Wörter für die Suche verwendet, dann ein Wort und schließlich drei Wörter. Je häufiger die Suchstrings genutzt wurden, desto besser ließen sie sich auch mit dem Standard Thesaurus Wirtschaft mappen (Neubert (2009)). Das mag daran liegen, dass kurze Suchanfragen häufiger genutzt wurden.

Bei der Analyse der EconStor Logs wurden ebenfalls die Logs aus 1,5 Jahren EconStor ausgewertet. Es zeigte sich, dass immerhin ein gutes Drittel der Suchen sich auf das kontrolliertes Vokabular abbilden ließ. EconStor bot bereits zu dem Zeitpunkt der Untersuchung eine Funktion zur Termerweiterung an, mit der die Suchenden ihre Suche um weitere automatisch vorgeschlagene Begriffe mittels ODER-Verknüpfung ausweiten konnten. Die Analyse der Logs zeigte, dass etwa 7 % der Nutzer von EconStor davon Gebrauch machten und in 1 % dieser Fälle dann besonders oft auf den Suchergebnisseiten weiterblättern (Borst (2012)).

Die erste Logfile-Analyse bezog sich nur anfragebasierte Daten. Bei der EconStor Studie wurde dagegen noch untersucht, ob nach einer Suchanfrage auch eine zweite oder folgende Ergebnisseite angesehen wurde. Eine Logfile-Analyse von EconBiz auf Basis der Sessions wurde an der ZBW bisher nicht durchgeführt.

3.4 EconDesk - Voruntersuchung

EconDesk ist der Online-Auskunftsservice der ZBW. Nutzer können hier per E-Mail Kontaktformular oder über ein Chat-Widget fragen an die ZBW stellen. EconDesk ist sowohl auf der Webseite der ZBW, als auch direkt in EconBiz eingebunden. Weiterhin findet man Verknüpfungen auf der ZBW Facebook-Seite und im OPAC Eonis. Der Chat wird in der Zeit von 9:00 Uhr bis 16:30 Uhr von Mitarbeitern betreut.

Im Vorfeld der Logfile Analyse von EconBiz wurden Anfragen an EconDesk ausgewertet. Die Vermutung lag nahe, dass diese Anfragen Hinweise auf generelle Probleme oder Bedürfnisse, die die Nutzer der ZBW beschäftigen (und die Nutzer von EconBiz) enthalten. Man muss dennoch festhalten, dass die Nutzung einer Online-Auskunft etwas anderes ist, als die Nutzung eines Recherchertools wie EconBiz. Nutzer müssen vermutlich auf größere Hürden treffen, ehe sie sich an EconDesk wenden. Viele weitere, die auch Probleme hatten, werden eventuell die Nutzung ganz abgebrochen haben.

Für die Auswertung wurden die jeweils initialen Anfragen der Nutzer an EconDesk in einer Zeitspanne von 10 Monaten im Zeitraum Oktober 2013 bis Juli 2014 zur Verfügung gestellt. Es handelte sich um 966 initiale Anfragen, die an EconDesk zu 80 % per E-Mail übersendet wurden.

Die Untersuchung zeigte, dass ein gutes Drittel der Anfragen bei EconDesk informationsorientiert sind, also im Wesentlichen nach Literatur oder Statistiken zu einem bestimmten Thema gesucht wurde. Mindestens ein Drittel der Anfragen wurde von Studierenden gestellt, was entweder explizit von den Personen in der Anfrage erwähnt wurde, oder durch die Nennung einer Aufgabe (z. B. „für meine Bachelorarbeit“) implizit zu erkennen war. Andere Zielgruppen ließen sich nur vereinzelt identifizieren. Gut ein Fünftel aller Anfragen gehörte zu den transaktionsorientierten Anfragen. Sie entstanden im Wesentlichen aus der Problematik auf ein bereits gefundenes Dokument zuzugreifen. Die Nutzer konnten nicht erkennen, ob und wie das Dokument für sie verfügbar war und baten um Hilfe oder die Zusendung eines PDFs des beschriebenen Dokuments. Hier scheint es also häufig zu Problemen für die Nutzer zu kommen.

Der folgende Abschnitt stellt die Methoden vor, die bei dieser Logfile-Analyse zum Einsatz kamen.

Kapitel 4

Methodik

4.1 Möglichkeiten und Grenzen der Logfile-Analyse

Ein Logfile ist keine genaue Aufzeichnung einer Suche. Es fehlen eigentlich die wichtigsten Elemente. Ein Logfile besteht im Wesentlichen aus URLs, die zu einem bestimmten Zeitpunkt von einem Server für einen Client abgerufen wurden. Dabei ist weder klar, wer der Client im Einzelnen ist, noch warum und wieso diese URL aufgerufen wurde. Ein etwas besseres Bild kann sich ergeben, wenn man gleich eine ganze Reihe von URLs hat, die in einen Zusammenhang und eine Reihenfolge gebracht werden können – eine sogenannte Session. Die einzelnen URLs erhalten so einen Kontext, aus dem man mehr Schlüsse zu ziehen hofft. Nicht alle Interaktionen müssen serverseitig aufgezeichnet worden sein, einiges kann auch im Cache beim Client gespeichert und von dort abgerufen worden sein (Yun 2009, 176). Diese Interaktionen fehlen dann im Logfile. Im Kontext von EconBiz dürfte dies jedoch nicht viele der für das Suchverhalten interessanten Seiten betreffen, da diese meist neu aufgerufen werden müssen (etwa wenn ein neue Suchanfrage formuliert wurde oder gefiltert wird).

Für die Qualität einer Logfile-Analyse von entscheidender Bedeutung ist die Art und Weise der Session-Rekonstruktion. Es gibt verschiedene Ansätze, wie es möglich ist Sessions zu rekonstruieren. Man gruppiert die Anfragen meist nach der IP-Adresse, von der sie ausgingen (server-seitig), oder nach Cookies (client-seitig) (Ozmutlu u. a. 2009, 348). Auch wenn man mit einiger Sicherheit davon ausgehen kann, dass zwischen den Transaktionen in einer Session ein auf diese Weise hergestellter Zusammenhang besteht, ist dieser keineswegs gesichert.

Weiterhin besteht das Problem, dass sich hinter einer IP-Adresse nicht zwangsläufig immer die gleichen oder auch nur einzelne Nutzer verbergen können. Selbst wenn man davon ausgehen könnte, dass die IP-Adressen sich nicht ändern (was sie tun), so ist nicht nachvollziehbar, wie viele Nutzer hinter einer Adresse stehen oder auch nur hinter einem Client (Yun 2009, 176).

Daneben enthalten Logfiles nicht nur die Request, die von „menschlichen“ Nutzern an den Server gestellt wurden. Eine Filterung nach automatischen Anfragen (Bots) ist also ebenso notwendig. Leider sind nicht alle Bots als solche gekennzeichnet, und so müssen eine

Reihe weiterer Ausschlusskriterien gefunden werden, die sich im Wesentlichen darauf stützen, dass man Phänomene im Logfile beobachtet, die ein „menschlicher“ Nutzer kaum auf „natürlichem Nutzungswege“ zustande bringen kann. Andererseits möchte man verhindern menschliche Nutzer herauszufiltern.

Zu guter Letzt ist eine Logfile-Analyse sehr begrenzt in ihrer Aussagefähigkeit, wenn es um Intentionen der Nutzer geht oder auch um die Frage nach der Intensität der Nutzung. Nur weil im Logfile steht, dass der Nutzer zwischen dem Aufrufen einer Suchergebnisseite und dem Anklicken eines Ergebnisses drei Minuten verstrichen sind, bedeutet dies nicht, dass der Nutzer sich in diesen drei Minuten die ganze Zeit mit der aufgerufenen Seite beschäftigt hat. Aus dem Logfile geht nicht hervor, mit welcher Aufmerksamkeit der Nutzer dem System gerade zugewandt war (Yun 2009, 172). Es wird jedoch oft davon ausgegangen, dass dies der Fall war.

Im Folgenden wird das Vorgehen der Sessionzuordnung bei dieser Logfile-Untersuchung dargestellt.

4.2 Datengrundlage

4.2.1 Der Logfile von EconBiz

Der Logfile von EconBiz verzeichnet die Transaktionen einzelner Tage von EconBiz und enthält sämtliche Anfragen, die an den EconBiz Server gestellt wurden. Aufgezeichnet werden die IP-Adresse des anfragenden Clients, die angefragte URL, der Zeitpunkt der Anfrage, der User Agent (verwendeter Browser), der zurückgegebene Servercode, der Typ der HTTP-Anfrage, die Abweichung der Zeitangabe von der General Mean Time und der Referer der gestellten Anfrage.

Neben den Anfragen für Suchergebnisseiten und einzelne Suchtreffer enthält der Logfile auch Ajax-Requests, über die beispielsweise die Informationen über die Verfügbarkeitsanzeige der Treffer angeboten werden. Außerdem enthält der Logfile die Anfragen für die statischen Elemente der EconBiz-Webseiten (Bildateien, wie z.B. das Logo etc.). Es fehlen jedoch Daten zum Ranking der Treffer, sowie Daten zu gefundenen Trefferzahlen bei Suchanfragen. Diese konnten im Rahmen der Untersuchung nur reproduziert werden, in dem die URLs noch einmal in EconBiz aufgerufen und die entsprechenden Daten notiert wurden.

Im Folgenden wird die Zuordnung der einzelnen Aufrufe zu Sessions beschrieben. Diese Zuordnung bildet das Herz einer Logfile-Analyse auf Sessionbasis, da hier die Zusammenhänge hergestellt werden.

4.2.2 Zuordnung von Sessions

Als Datengrundlage für die Auswertung dienten die Logfile-Daten aus den Tagen vom 30. November 2013 bis zum 30. November 2014, also genau ein Jahr an Logfile-Daten. Auf diese Weise können auch semester- und vorlesungsbedingte Nutzungsschwankungen ausgeglichen werden.

Aus der Menge der Logfile – Daten Sessions zu formen, von denen man mit einiger Sicherheit behaupten kann, dass menschliche Nutzer dahinter stehen ist nicht trivial. Jansen et al. haben ermittelt, dass die sicherste Methode eine Mischung aus IP-Adressen, Cookies und Suchanfrageinhalten ist (Jansen u. a. 2007a, 870). Letztere dienen hier als Ersatz für einen zeitlichen Cut-Off (Jansen u. a. 2007a, 866). Bei dieser Methode werden Sessions getrennt, wenn sich die Inhalte der Suchanfragen komplett verändern und nur noch auf einem höheren (semantischen) „Abstraktionslevel“ ein Zusammenhang besteht. Da es bei Suchtaktiken genau hierum geht, erschien diese Methode für diese Untersuchung nur bedingt sinnvoll. Auch standen Cookies für diese Untersuchung nicht zur Verfügung.

Die Zuordnung der Sessions wurde für diese Untersuchung aufgrund der IP-Adresse und des User Agents (verwendeter Browser) gemacht. Die gleiche IP-Adresse mit unterschiedlichen Browsertypen wurde also in zwei verschiedene Sessions eingeordnet. Daneben wurde ein Cut-Off von 30 Minuten gewählt. Kam keine weitere Transaktion von der gleichen IP mit dem gleichen Browsertyp innerhalb der folgenden 30 Minuten wurde die Session beendet. Später erfolgte Transaktionen wären also Teile einer neuen Session.

Herausgefiltert wurden die Nutzungsdaten von ZBW-internen IP-Adressen. In den meisten Fällen handelte es sich um Mitarbeiterzugriffe auf EconBiz. In den Lesesälen der ZBW wird EconBiz nicht direkt beworben. Die Lesesäle ließen sich anhand einer einzelnen IP-Adresse identifizieren. Da jedoch bekannt ist, dass hinter dieser IP-Adresse mehrere verschiedene Nutzer gleichzeitig stehen könnten, wurde auch diese Lesesaaladresse aus der Untersuchung herausgefiltert.

4.2.3 Datenreinigung

Nach dieser Sessionzuordnung und dem Herausnehmen der offensichtlich automatisch generierten Logfile-Daten (Google Bot etc.), musste weiter gefiltert werden, da noch immer ein beträchtlicher Anteil automatischer Daten vermutet werden konnte. Als Benchmark dienten die Zahlen aus dem eTracker-Tool, das die ZBW für WebAnalytics in EconBiz verwendet. Es wurde probeweise der Bot-Filter und die Sessionzuordnung für einen einzelnen Tag vorgenommen und diese Zahlen mit den eTracker Daten zu einzelnen Besuchern und Seitenaufrufen

verglichen und passende Kriterien gesucht, nach deren Anwendung die Daten aus der Logfile-Datenbank mit denen aus dem eTracker gut übereinstimmten.

In der Literatur zur Untersuchung der Logfiles wird oft ein Cutoff der Sessionlänge gesetzt (oft bei 100 Anfragen pro Session) und davon ausgegangen, dass oberhalb dieser Grenze nichts mehr von menschlichen Nutzern stattfindet, jedoch gleichzeitig in Kauf genommen, dass eine Reihe von automatisch generierten Sessions noch im Datensatz sind (Jansen u. a. 2007b, 332). Ein anderer Ansatz wäre die Beobachtung von bestimmten Navigationsmustern (Tan u. Kumar (2004)), etwa eine sehr hohe Seitenauffrequenz. Bei der Durchsicht längerer Sessions in EconBiz entstand jedoch oft der Eindruck, dass es sich hier kaum um automatische Sessions handelte, da oft sehr vernünftige Reformulierungen der Suchanfragen erfolgten.

In dieser Untersuchung konnte ein besseres Merkmal zur Unterscheidung gefunden werden. Von menschlichen Nutzern im Browser aufgerufenen Anfragen benötigen auch Ajax-Abfragen, die z.B. die Anzeige über die Verfügbarkeit der Dokumente nachladen. Fehlten solche Ajax-Abfragen in den Sessions, konnten diese als automatisch generiert herausgefiltert werden. Des Weiteren gab es eine Reihe von „kaputten“ Sessions, bei denen die erste Transaktion bereits EconBiz als Referer angegeben hatte. Dies könnte unter anderem durch den 30 Minuten Cutoff entstanden sein, oder durch einen automatischen IP-Wechseln innerhalb einer Session. Auch diese Sessions wurden aus der Untersuchung herausgenommen. Die Zahlen stimmten nach diesen Maßnahmen ziemlich genau mit den eTracker-Daten überein (mit etwa 1 % Abweichung).

Nach dem Herausfiltern der gekennzeichneten Bots, der Sessions von den ZBW-IPs wurden etwa 6 Millionen Sessions mit ca. 13 Millionen Transaktionen in die Datenbank übernommen. Nach Herausfiltern der Sessions ohne Ajax-Abfragen, ohne mindestens einen Seitenaufruf und des Filters für „kaputte“ Sessions, blieben noch 3.463.823 Mio. Sessions für den Untersuchungszeitraum.

4.2.4 Gründe für die Aussortierung von Daten

Um einen Überblick zu bekommen, welches der Aussortierungskriterien für die Datenbereinigung griff und wie viele Session von den einzelnen Kriterien betroffen waren, wurden die aussortierten Sessions kurz betrachtet.

Die Datenbank wurde aufgrund von drei Kriterien von Sessions bereinigt. Es wurden Sessions ausgeschlossen, die die Länge 0 hatten (keine „echten“ Seitenaufrufe z.B. nur Javascript Anfragen), die keine Ajax-Requests ausgelöst haben (Hinweis auf automatischen Abruf), oder die als Referer der ersten Aktion in der Session bereits „EconBiz“ angegeben hatten.

	Sessionlänge 0	Ohne Ajax-Requests
Ohne Ajax-Requests	0,44% (9.048)	-
Mit Referer EconBiz	3,25 % (66.706)	13,3 % (272.348)

Tab. 4.1: Gründe für die Aussortierung beim Bot Filtern

Theoretisch dürfte der Referer der ersten Aktion nicht EconBiz sein. Wenn er tatsächlich bereits vorher auf EconBiz war, war er entweder sehr lange inaktiv und es kann nicht mehr von der gleichen Session gesprochen werden, oder es liegt ein anders gearteter Fehler vor. Die Auswertung der aussortierten Sessions betraf insgesamt 2.048.019 Sessions (35,44 % aller Sessions). Tabelle 4.1 zeigt die Aufstellung der Verteilung der Ausschlusskriterien unter den Sessions.

Das Kriterium der Sessionlänge traf auf 8,97 % (183.788) Sessions zu. Dass keine Ajax-Requests vorkamen, war in 91,59 % (1.875.752) der Fälle, und dass die Sessions auch kaputt waren, traf in 22,05% (451.506) der Fälle zu. Alle drei Kriterien waren nur in 0,93 % (19.033) Fällen vorhanden.

Bei 72,24 % (1.479.431) war allein das Fehlen von Ajax-Requests der Grund für den Ausschluss. 3,63 % (74.386) der Sessions wurden allein wegen des Referers der ersten Transaktion ausgeschlossen. 0,34 % (6.926) Sessions wurden allein aufgrund der Länge 0 ausgeschlossen. Das Hauptkriterium für den Ausschluss der Daten war also das Fehlen von Ajax-Requests in den Sessions. Eine Reihe weiterer Sessions sind leider bei der Sessionzuordnung auseinander gefallen. Es kann also sein, dass unter den noch den Untersuchungsdaten verbliebenen Sessions die zweite Hälfte fehlt, oder aus der Mitte etwas herausgefallen ist. Geht man davon aus, dass die kaputten Session jeweils die zweite Hälfte einer in der Menge verbleibenden Session (oder ein Mittelteil daraus) darstellen, so dürften maximal 2,1 % der in der Datengrundlage verbliebenen Sessions davon betroffen sein.

4.3 Stichprobenbildung

Für die Fragestellungen, die eine intellektuelle Interpretation erfordern ist es jedoch kaum möglich, die Gesamtheit der Sessions durchzugehen. Es musste also zu einer kleineren Stichprobe gekommen werden.

Eine einfache Zufallsauswahl aus der Gesamtheit zu treffen erschien jedoch in mehrerer Hinsicht wenig sinnvoll. So ist nach einer ersten Durchsicht der Daten schnell erkennbar gewesen, dass ein Großteil der Sessions nur eine einzige Seite aufgerufen hatte (70 %), noch dazu in den meisten Fällen einen Einzeltreffer. 70 % der Sessions sind damit eigentlich zunächst un-

interessant für Fragestellungen, die auf Reformulierungstaktiken von Suchanfragen und die Unterscheidung von Known-Item-Suche und thematischer Suche abzielen.

Aus diesen Gründen wurde eine andere Grundgesamtheit für die Stichprobenziehung gebildet, enthaltend Sessions mit folgenden Eigenschaften:

1. In der Session wird mehr als eine Seite aufgerufen.
2. In der Session wird mindestens eine Suchergebnisseite aufgerufen.
3. Sessions, die den Referer „zbw.eu“ haben, können auch nur einen Seitenaufruf enthalten, wenn diese Seite nicht die Homepage, sondern eine Suchergebnisseite ist (hier wurde der EconBiz-Suchschlitz auf der Homepage der ZBW verwendet).

Es verblieben in der Untersuchungsmenge noch 459.097 Sessions (etwa 8.000 davon durch die Ausnahmeregelung mit dem Suchschlitz auf der ZBW-Homepage). Aus diesen Sessions wurde zufällig für jeden Tag im untersuchten Jahr fünf Sessions für die intellektuelle Interpretation ausgewählt. In der Stichprobe befanden sich also 1.825 Sessions, gleichmäßig über das gesamte Jahr verteilt, die länger sind, als der Großteil der Sessions und damit nur für einen (interessanten) Anteil der Gesamtdaten repräsentativ sein können.

4.4 Datenerfassung

Die Datenerfassung erfolgte in einer MySQL-Datenbank. Die Datenbank bestand im wesentlichen aus zwei Tabellen: Daten zu der entsprechenden Session insgesamt (Seitenaufrufe, Dauer, Herkunft, Zugehörigkeit zur kleineren Stichprobe) und Daten zu den einzelnen Transaktionen in der Session. Die Tabellen konnten mittels eines eindeutigen Identifiers für jede Session verknüpft werden. Daneben gab es eine eigene Tabelle für die Suchanfragen und eine Tabelle für die Erfassung der in den Suchanfragen enthaltenen Elemente. In der großen Tabelle mit den Transaktionen waren zudem die entsprechenden Felder vorhanden, um etwa zu vermerken, ob es sich bei der Aktion um eine Known-Item-Suche handelte, welche Taktik gerade zum Einsatz kam und um was für eine Seite es sich eigentlich handelte, die dann aufgerufen wurde. Neben diesen Daten wurde also erfasst, um was für einen Seitentyp es sich handelt: den Objekttyp.

4.4.1 Objekte

Um die Abläufe im Logfile zu erfassen, wurde notiert, was für ein Objekttyp die Seite ist. Da nicht nur Suchergebnisseiten, sondern auch alle anderen EconBiz-Seiten im Logfile verzeichnet sind, kann man erkennen, welche Einzeltreffer die Nutzer angesehen haben. Es wurden

verschiedene Objekt-Typen vergeben, die meist aufgrund der URL-Struktur automatisch zugeordnet werden konnten:

Einzeltrefferseite: Einzelansichten der in EconBiz verzeichneten Treffer.

Suchergebnisseite: darunter auch Suchergebnisseiten aus der erweiterten Suche und Suchergebnisseiten aus der Suche nach „ähnlichen Treffern“.

Homepage: die Startseite von EconBiz.

CMS: Seiten aus dem umgebenden Typo3, wie etwa „Über Econbiz“ etc.

Bookmarks: Seiten, die unter der Merklisten-Funktion liegen.

Login: Aufruf der Login-Seite und deren Unterseiten.

EventResults: Seiten, die in der Veranstaltungs-Suche aufgerufen wurden.

4.4.2 Aktionen

Neben den Objekten wurde als weiteres Datum eine „Aktion“ erfasst, um darzustellen, was der Nutzer zwischen den beiden Seitenaufrufen wahrscheinlich getan haben wird. Sie sind anders als die Taktiken, nicht dafür da die Intentionen der Nutzer zu erfassen, sondern Abläufe zu beschreiben. Sie helfen zu unterscheiden, auf welche Weise eine Taktik angewandt wurde. In Kombination mit den vergebenen Strategien (s. 6.1) beschreiben sie ziemlich genau, was in der Session vorgefallen ist. Hier eine Liste mit den vergebenen Aktionen, die ausgeführt werden konnten:

Submit: Das Abschicken einer Suchanfrage über einen Suchschlitz entweder der Standard- oder der Erweiterten Suche.

Select: Das Auswählen eines Treffers auf einer Suchergebnisseite, auch das Auswählen eines Treffers in einer anderen Suchmaschine, als Einstieg in EconBiz.

Return: Das Zurückgehen auf eine vorher aufgerufene Seite.

Scroll: Das Weiterblättern auf den Suchergebnisseiten in EconBiz.

Save: Das Speichern einer Datei über EconBiz, etwa der Aufruf eines PDFs aus dem Archiv der USB Köln.

Resubmit Erneutes Abschicken einer bereits gesuchten Suchanfrage.

Bookmark Das Speichern eines Einzeltreffers oder einer Suchergebnisseite in den Merklisten.

Chain: Das Klicken auf einen der Links, die eine Suche nach dem entsprechenden Feld in EconBiz auslöst (z.B. alle Artikel der entsprechenden Zeitschrift). Meist in Verbindung mit der entsprechenden Taktik, die genauer beschreibt, was für ein Link geklickt wurde.

Cite: Das Klicken auf das Zitierangebot bei Einzeltreffern

Export: Das Klicken auf das Exportierangebot der Literaturangaben eines Einzeltreffers.

Email: Das Klicken auf das Email-Icon auf einer Einzeltreffer-Seite.

4.4.3 Erfolgsbewertung

Um den Sucherfolg in einer Logfile-Analyse bewerten zu können, müssen Indikatoren definiert werden, da die Nutzer selbst über ihre Erfolgsbewertung und Zufriedenheit mit den gefundenen Ergebnissen nicht befragt werden können. Ein häufig verwendeter Indikator ist das Klickverhalten (z.B. Huang u. Efthimiadis (2009)). Ein Klick auf einen Treffer (Clickthrough) wird als Interessenbekundung durch den Nutzer gewertet. Neben den Klicks kann auch die Retrievalleistung für die Messung der Systemperformance betrachtet werden. Es kann also auch das Ranking als Erfolgsindikator herangezogen werden (z.B. Lewandowski (2011)).

Bei der Bewertung des Erfolgs aufgrund von Clickthrough-Daten ergab sich in dieser Untersuchung ein grundlegendes Problem: Wenn im Logfile nicht erkennbar ein Treffer angeklickt wurde, bedeutete das leider nicht, dass auch kein Klick erfolgt ist. Die Klicks auf die Verfügbarkeitsprüfung in der Trefferliste beispielsweise führen den Nutzer zu einem Dienst des hbz Nordrhein-Westfalen und somit weg vom EconBiz-Logfile. Gerade der Klick auf die Verfügbarkeitsprüfung wäre jedoch ein guter Indikator für das inhaltliche Interesse des Nutzers an dem Treffer. Kein Klick bedeutet also in dieser Untersuchung nicht zwangsläufig Misserfolg. Des Weiteren mussten die Daten für die Trefferzahlen und das Ranking einzelner Treffer nachträglich ermittelt werden. Hierzu wurden die Suchergebnisseiten während der intellektuellen Analyse erneut aufgerufen und Trefferanzahl und Ranking angeklickter Einzeltreffer (oder gesuchter Known-Items) notiert. Da die Reproduzierung der Ergebnisse jedoch bis zu einem Jahr später erfolgt ist, als die Suche des Nutzers, und der Index sich in der Zwischenzeit verändert hat, ist diese Messung alles andere als exakt. Es kann also oft angezweifelt werden, ob sich das Suchergebnis für den Nutzer in ähnlicher Weise dargestellt hat, als die Original-Session stattfand. Für Fälle in denen evident das Suchergebnis nicht mehr auf der richtigen

Suchergebnisseite angezeigt wurde (etwa auf Seite 2, ohne dass der Nutzer weitergeblättert hatte), wurde Rang 0 vergeben (also keine Angabe).

Da es bei Known-Item-Suchen möglich war, ein „richtiges“ Ergebnis zu erzielen, wurden die Bewertungen von Erfolg oder Misserfolg in Abhängigkeit vom Suchtypus bewertet.

Bei thematischen Suchen ließ sich der Erfolg nur bemessen, wenn ein Klick auf einen Einzeltreffer erkennbar war und der Misserfolg nur im Falle von Nulltreffer-Ergebnissen sicher festzustellen. Hierbei wird angenommen, dass der Zuwachs im Index größer ist, als das Herausnehmen von Daten und daher die Wahrscheinlichkeit, dass eine Nulltreffer-Seite bei der Original-Session noch Treffer hatte und bei der Reproduzierung nicht mehr, sehr klein ist.

Für Known-Item-Suchen sinnvolle Maße von Retrievaleffektivität sind Success@n und Mean Reciprocal Rank (Lewandowski 2011, 214). Success@n betrachtet für eine Menge von Suchanfragen, die Anteile auf welchem Rang jeweils das erste „richtige“ Ergebnis zu finden war (Lewandowski 2011, 357). Der Mean Reciprocal Rank gibt an einen durchschnittlichen relativen Rang der gesuchten Items an. Die Ränge werden dabei unterschiedlich gewichtet. Meist erfolgt hinter Rang 5 keine Gewichtung mehr (es werden 0 Punkte vergeben). Ein ideales Ergebnis wäre hier 1 (das Known-Item war stets auf Rang 1) (MacFarlane 2007, 354).

Daneben wurde auch im Falle von Known-Item-Suchen dichotom ein Erfolgzustand bewertet. Unter folgenden Umständen wurde ein Erfolg der Suche gewertet:

1. Das gesuchte Item wurde vom Nutzer im Logfile erkennbar angeklickt (der Einzeltreffer wurde ausgewählt)
2. Das gesuchte Item wurde zwar nicht im Logfile erkennbar angeklickt, bei der Reproduzierung der Suche kam das gesuchte Item jedoch auf Rang 1 der Ergebnisliste.

Da es bei dieser Forschungsfrage auch um eine Bewertung der Retrievalleistung von EconBiz geht, war für diesen Fall das Anklicken des gesuchten Items nicht allein ausschlaggebend für den Erfolg. War das Known-Item auf der ersten Seite zu finden, wurde der Rang notiert. War es nicht zu finden, wurde Rang 0 vergeben, sofern es überhaupt in EconBiz indexiert war. Bei nicht indexierten Known-Items wurde kein Rang notiert.

Suchte der Nutzer nur nach einem Nachnamen, und gibt es mehrere passende Entsprechungen in EconBiz, der Nutzer wählte aber nicht erkennbar einen Treffer aus, so wurde nicht davon ausgegangen, dass die Suche ein Misserfolg war. In diesen Fällen wurde ein Erfolg gewertet, wenn der erste Treffer der Suchergebnisseite bei der Reproduzierung einen Autor des gesuchten Namens enthielt.

Klickt der Nutzer hingegen auf ein Item, und stimmen die Angaben aus der Suchanfrage mit den Daten des Einzeltreffers überein, so wurde davon ausgegangen, dass der Nutzer auch das angeklickte Item finden wollte.

Ein Problemfall war das Auftreten von Information Encountering-Effekten. Diese waren in der Datenstruktur so nicht vorgesehen. Es kam dennoch vor, dass man anhand der Reformulierungen des Nutzers ein Known-Item identifizieren konnte, der sichtbare Klick am Ende jedoch auf ein völlig anderes Ergebnis gesetzt wurde. Mit einem Klick scheint die Suche erfolgreich, aber es wurde nicht das Known-Item geklickt, das eigentlich gesucht wurde (oft weil dieses gar nicht in EconBiz auffindbar war). Ähnliche Verzerrung hatte auch das Dubletten-Problem auf die Erfassung der Retrievalleistung: Wenn das Known-Item auf den ersten drei Rängen gleichzeitig zu finden war, hat der Nutzer oft nicht das erste ausgewählt (vorausgesetzt die Reproduzierung ist gleich mit der Situation in der Originalsession). Da jedoch nur ein Datenfeld für die Erfassung vorgesehen war, wurde der Klick über die Retrievalleistung gestellt, da es in erster Linie um das Suchverhalten der Nutzer ging.

Als Misserfolg einer Known-Item-Suche wurde bewertet, wenn die Suche in der Reproduzierung 0 Treffer erzielte, oder das gesuchte Item weder angeklickt noch auf Rang 1 gerankt wurde.

Kapitel 5

Allgemeine Ergebnisse

5.1 Suchanfragen in EconBiz

5.1.1 Häufigkeit des Vorkommens

Um ein Bild davon zubekommen, welche Anfragen am häufigsten gestellt werden, wurde einmal über die Gesamtheit der Daten ausgewertet, zunächst pro Transaktion. Diese Zählung enthält also auch solche Anfragen, die durch Zurückgehen oder Weiterblättern mehrmals in der Session auftauchen, ebenso wie evtl. technisch bedingte Reloads.

Bedingungen der Abfrage waren:

1. Die Transaktion enthält eine Suchanfrage.
2. Die Transaktion hat entweder den Referer Econbiz oder zbw.eu (keine Suchmaschine).
3. Die URL der Transaktion enthält nicht den Zusatz „limit=20“ (Hinweis auf automatisch generierte Suchanfrage).
4. Es wurde im Feld „AllFields“ gesucht.
5. Die Nummer der Suchergebnisseite ist 1 (kein Weiterblättern).

Die Bedingung der Feldbeschränkung wurde noch einmal nachträglich hinzugefügt, da sich zeigte, dass besonders bei den Suchanfragen mit großer Anzahl oft eine Feldbeschränkung (z.B. „Subject“) vergeben war, was wieder ein Hinweis darauf ist, dass der Nutzer auf ein Schlagwort eines Einzeltreffers geklickt und so eine Suche ausgelöst hat. Es blieben noch 281.535 verschiedene Suchanfragen. Auch hier sind noch automatisch generierte enthalten, „isPartof:“ oder „type_genre“, die aus der Liste gelöscht und für die weitere Analyse ignoriert wurden, was die Zahl auf 278.447 unterschiedliche Anfragen reduziert.

Diese Menge Suchanfragen enthält noch immer Reloads und Zurückgeh-Aktionen, und eventuell sind einige echte Suchanfragen durch die Hinzunahme der Feldbeschränkung herausgefallen, dafür enthält das Ergebnis allerdings weniger Ballast durch Klicks auf Schlagworte.

Die häufigsten zehn Abfragen 2014 waren demnach betrachtet über die Transaktionen:

1. Marketing (707)
2. Crowdfunding (530)
3. Controlling (399)
4. Employer branding (375)
5. Change Management (340)
6. Balanced scorecard (340)
7. Risikomanagement (296)
8. Unternehmensbewertung (287)
9. Generation Y (286)
10. Mindestlohn (272)

Die gleiche Abfrage mit den gleichen Beschränkungen ausgeführt über die unterschiedlichen Sessions ergibt folgendes Bild:

1. Marketing (355)
2. Controlling (213)
3. Crowdfunding (200)
4. Employer branding (170)
5. Change management (150)
6. Risikomanagement (143)
7. Mindestlohn (135)
8. Social Media (131)
9. Nachhaltigkeit (127)
10. Unternehmensbewertung (120)

91,23 % der Anfragen kommen nur in einer Session vor, d.h. nur 8,77 % der Anfragen kommen in mehr als einer Session vor.

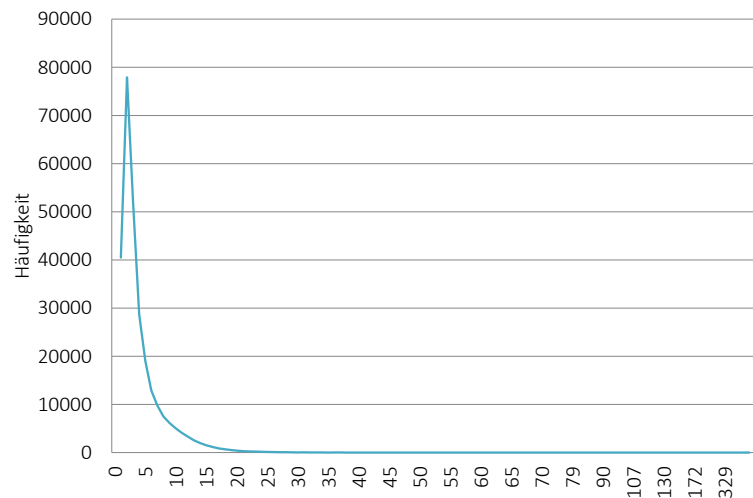


Abb. 5.1: Absolute Häufigkeiten der Länge von Suchanfragen in Worten

5.1.2 Länge der Suchanfragen

Betrachtet über die Anfragen aus der oben beschriebenen Abfrage, bestehen 14,48 % aus nur einem Wort, 27,94 % aus zwei Worten und noch weitere 18,63 % aus drei, sowie 10,23 % aus vier Worten. Ab 13 Worten fällt die Häufigkeit unter 1 %. Bereits 61,25 % der Anfragen enthalten bis zu drei Worte, 91,39 % bis zu neun Worte und 99,01 % bis zu 18 Worte. Der Mittelwert liegt bei 4,2 Worten pro Anfrage. Der Median liegt zwischen 2 und 3 Worten. Das Maximum lag bei 931 Worten, ein Fall, wo ein kompletter Absatz in den Suchschlitz kopiert wurde.

Die Ergebnisse dieser Analyse sind sehr ähnlich den Ergebnissen aus der Logfile-Analyse von Econis aus dem Jahr 2009 (s. 3.3). Auch hier stellte sich heraus, dass die meisten Suchanfragen aus zwei Worten bestehen, dann drei und dann eins. Sogar das beliebteste Suchwort „Marketing“ ist gleich. Dagegen ist „Mindestlohn“ wahrscheinlich ein Saison-Suchwort, weil es ein aktuelles Thema von 2014 war.

5.2 Sessiondauer

Die durchschnittliche zeitliche Länge einer Session betrug 1 Stunde 52 Minuten und 14 s. Über die gesamte Datenmenge konnte der Median durch die Menge der Daten nicht ermittelt werden. Vermutlich ist der Durchschnitt jedoch stark verzerrt durch Ausreißer. Es wurde im Schnitt alle 91,89 s eine Seite aufgerufen. Die durchschnittliche zeitliche Dauer von der

ersten bis zur letzten Transaktion in den Sessions der Stichprobe betrug 6,46 Minuten. Der Median liegt bei 1 Minute und 47 s. Es wurde im Schnitt alle 57 s eine Seite aufgerufen. Die Verweildauer auf der letzten aufgerufenen Seite wird naturgemäß nicht im Logfile erfasst. Ebenso fehlen die Zahlen für den großen Teil der Sessions, die nur eine einzelne Seite aufrufen.

5.2.1 Verweildauer auf Suchergebnisseiten

Die Verweildauer auf Suchergebnisseiten kann nur für solche Suchergebnisseiten errechnet werden, denen eine weitere Aktion in der Session folgte, da ansonsten die zeitliche Differenz fehlt. Die durchschnittliche Verweildauer auf einer Suchergebnisseite in der Stichprobe betrug 55,97 s, wobei die Verteilung stark linksschief ist. Der Median liegt bei 16 s. Die durchschnittliche Verweildauer auf Suchergebnisseiten mit 0 Treffern lag bei 38,9 s. Die Verteilung ist auch stark linksschief. Der Median liegt bei 8 s. Auf Suchergebnisseiten, auf denen ein Treffer ausgewählt wurde, verweilten die Nutzer im Schnitt 53,94 s. Auch hier ist die Verteilung stark linksschief, der Median liegt bei 18 s. Da kein sichtbarer Klick nicht bedeutet, dass keine Auswahl stattfand, können die Verweildauern auf den Suchergebnisseiten mit und ohne Treffer Auswahl leider nicht verlässlich verglichen werden. Die Durchschnitte deuten jedoch keinen erheblichen Unterschied an (53 s bei Klick gegenüber 55 s ohne Klick). Dies könnte aber eben auch durch nicht beobachtete Klicks entstanden sein.

5.3 Bounce-Sessions: Sessions mit nur einem Seitenaufruf

Um zu untersuchen, was in den Sessions passiert ist, die nicht in die Untersuchungsmenge gefallen sind, wurden unter anderem die Sessions untersucht, bei denen die Nutzer nur einmal (und dann mind. 30 Minuten lang nicht mehr) in EconBiz eine Seite aufgerufen hatten. Dies betraf 2.449.689 Sessions, also 70,72 % der Nutzersessions im Untersuchungszeitraum. Beim überwiegenden Anteil dieser Sessions wurde nur eine Einzeltrefferseite aufgerufen (65 % der „Bounce“-Sessions), in weiteren 4,86 % wurde eine Suchergebnisseite aufgerufen, noch 0,44 % haben sich eine Einzeltrefferseite unter der Eventsuche angesehen. Alle weiteren Seitentypen liegen unter 0,1 %.

Je nachdem, auf welchem Seitentyp die Nutzer ankommen, unterscheidet sich das Abbruchverhalten. 77 % der Nutzer, die auf einem Einzeltreffer ankommen (aus der gesamten Datenmenge) rufen keine weitere Seite in EconBiz auf. 55 % der Nutzer, die auf einer Suchergebnisseite ankommen, rufen keine weitere Seite in EconBiz auf und lediglich 1 % der Nutzer, die die Homepage aufgerufen haben, rufen keine weitere Seite in EconBiz auf.

Im überwiegenden Teil wurden also durch Suchmaschinen direkt auf Einzeltreffer in EconBiz gesprungen. Sie machen einen Großteil der Datenmenge aus, sagen jedoch leider gar nicht viel über das Suchverhalten in EconBiz, außer, dass die meisten Nutzer wieder abspringen, weil sie entweder gefunden haben, was sie suchen, oder nicht motiviert sind, eine Suche in EconBiz aufzunehmen. Es gibt jedoch noch eine weitere Menge Sessions, in denen eigentlich kaum ein Verhalten in EconBiz gemessen werden kann, wie der folgende Abschnitt zeigt.

5.4 Rebounding

Mit „Rebounding“ ist das Verhalten gemeint, mehrmals in einer Session von einer Suchmaschine aus, mitten in EconBiz hineinzuspringen. In 13,04 % (238) der Sessions in der Stichprobe wurde mehrmals von einer Suchmaschine direkt auf eine EconBiz-Seite gegangen. Im Mittel wurden nur 47 % der Seiten in den Rebounde-Sessions nicht direkt von einer Suchmaschine aus aufgerufen. D.h. das mindestens die Hälfte der Seiten in diesen Sessions in Suchmaschinen ausgewählt wurden, statt in EconBiz. Im Mittel wird 2,5 Mal von einer Suchmaschine in EconBiz hineingesprungen. Auch der Median liegt bei 2.

Ein Versuch sich der Zahl der Rebounder über die gesamte Datenmenge anzunähern, wurde unternommen, indem die Menge der unterschiedlichen URLs in den Sessions gezählt wurden, bei denen der Referer „google“ „yahoo“ oder „bing“ lautete. Dies blendet zwar eine Vielzahl weiterer Suchmaschinen aus, die jedoch meist nur vereinzelt vorkommen. Gleichzeitig wird das bloße Zurückgehen auf dieselbe Seite aus der Zählung herausgefiltert.

In 196.257 Sessions wurde wahrscheinlich mehrmals von einer Suchmaschine direkt in EconBiz gesprungen. Das entspricht 5,66 % aller Sessions. Lässt man jedoch die Bounce-Sessions (Länge 1) heraus, sind es schon 19,35 % also fast ein Fünftel der längeren Sessions, in denen wahrscheinlich Rebounding vorkam. Auch hier wurde im Mittel 2,2 Mal von einer Suchmaschine aus in EconBiz gesprungen und auch hier liegt der Median bei 2 Hereinsprüngen pro Session.

Das Verhältnis zwischen durch Hereinspringen aufgerufenen Seiten und in EconBiz aufgerufenen Seiten ist in dieser Menge noch schlechter, als in der Stichprobe. Nur 22 % der Seiten waren durchschnittlich nicht in einer Suchmaschine ausgewählt. Der Median liegt sogar bei 0. In 56,53 % (110.961) dieser Sessions wurden die Seiten nur in Suchmaschinen ausgewählt und in EconBiz gar nichts weiter angeklickt oder gesucht.

Zu den 70,7 % der Sessions, in denen nur eine einzige Seite aufgerufen wurde, müssen also noch die hinzugezählt werden, in denen eigentlich auch nur in einer anderen Suchmaschinen Treffer ausgewählt wurden, die dann nach EconBiz geführt haben, so dass man von min-

destens 73,93 % der Sessions nicht sagen kann, ob überhaupt eine Aktion in EconBiz oder durch EconBiz stattgefunden hat.

Die zeitliche Frequenz des Seitenaufrufs in Rebound-Sessions ist in der Stichprobe größer, als über alle Sessions gesehen. Das Mittel liegt bei einem Seitenaufruf alle 119 s (Median 76 s). Die Zeitabstände der Seitenaufrufe sind also im Durchschnitt mehr als doppelt so groß, wie in nicht Rebound-Sessions. In Sessions, in denen maximal einmal von einer Suchmaschine aus hineingesprungen wurde, lag der Durchschnitt bei einem Seitenaufruf alle 48 Sekunden (Median 21 s) ($t=2,53 \times 10^{-15} < 0,05$).

5.5 Sichtbare Klicks und Trefferzahlen

Um einen Eindruck zu bekommen, wie oft ein Klick im Logfile erkennbar war, wurde über alle über den Suchschlitz gestellten Suchanfragen in der Stichprobe Trefferzahlen, Suchtyp, Rank des geklickten (oder gesuchten) Treffers und Sucherfolg ausgewertet. Gewertet wurden zunächst nur die ersten Suchergebnisseiten (ohne weiterblättern).

Über alle Suchergebnisseiten wurde nur in 43,21 % ein Rang für einen Treffer notiert. Ein Erfolg wurde in 32,31 % der Suchanfragen vergeben. Die Zahl der sichtbaren Klicks reduziert sich, wenn man bedenkt, dass bei Known-Item-Suchen bereits das Ranking auf Rang 1 ausgereicht hat, um einen Erfolg zu registrieren. Rechnet man diese Fälle heraus liegt die Quote sichtbarer Klicks bei nur noch 20,6 % auf den ersten Suchergebnisseiten nach dem Abschicken einer Suchanfrage. Dabei unterscheiden sich die Suchtypen nicht. Der Anteil sichtbarer Klicks auf den ersten Suchergebnisseiten liegt bei thematischen und Known-Item-Suchen bei 20 % bzw. 21 %.

In dieser Untersuchung wurden also entweder eine Menge Erfolgsindikatoren bei der Suche übersehen, weil die Klicks auf die Verfügbarkeitsprüfung ging und damit nicht im Logfile erkennbar war, oder es wurde tatsächlich sehr wenig ausgewählt in EconBiz. Die Zahlen zum Sucherfolg sollten jedenfalls mit einiger Vorsicht betrachtet werden, insbesondere bei thematischen Suchen.

Die durchschnittliche Trefferanzahl nach über den Suchschlitz gestellten Suchanfragen liegt bei 140.256 Treffern, der Median bei 29 Treffern. Die Verteilung ist stark linkschief. Die Nulltreffer-Quote liegt bei 19,64 % der Suchergebnisseiten nach einer Suchanfrage in der Stichprobe. Der Anteil ist bei Known-Item-Suchen jedoch deutlich höher mit 25,73 %.

5.6 Weiterblättern

Auf Suchergebnisseiten weitergeblättert wurde nur in 16,77 % der Sessions in der Stichprobe. Im Schnitt wurde dann 3,7 Mal geblättert. D.h. wenn geblättert wurde, dann gleich mehrmals in der Session. Wenn weitergeblättert wurde, wurde nur in 5 % der Transaktionen eine Suchergebnisseite über Seite 19 aufgerufen. 62,14 % riefen maximal Seite 3 auf und nur 23,21 % riefen noch Seite 7 auf. Auch diese Zahlen sind etwas mit Vorsicht zu sehen, da durch den Einstieg über andere Suchmaschinen manche Nutzer auf höheren Suchergebnisseiten ankamen, als auf Seite 1. Nur in 13,95 % der Weiterblätter-Aktionen in der Stichprobe konnte ein Klick auf ein Suchergebnis erkannt werden. Das könnte darauf hindeuten, dass die Nutzer beim Weiterblättern seltener ein Treffer anklicken, als auf der ersten Suchergebnisseite. Vielleicht nutzen sie aber auch die Verfügbarkeitsprüfung.

Rechnet man Suchergebnisseiten aus der Gesamtmenge der Daten heraus, die aus anderen Suchmaschinen aufgerufen wurden, und zählt in wie vielen Sessions eine Suchergebnisseite größer 1 aufgerufen wurde, so wurde nur in 2,02 % der Sessions mindestens einmal wahrscheinlich weitergeblättert. Das Verhältnis zwischen Sessions und Transaktionen beträgt bei dieser Betrachtung 4,37. Auch hier wurde also wahrscheinlich mehrmals weitergeblättert, wenn geblättert wurde. Die etwas höhere Zahl erklärt sich vielleicht durch Zurückgeh-Aktionen zwischen Suchergebnisseiten.

5.7 Ajax-Aktionen

Über die Aktionsleiste auf der Einzeltrefferseite kann man Zitierdaten exportieren oder sich per Mail schicken. Auch auf Suchergebnisseiten können Suchergebnisse in einer Merkliste gespeichert werden. Auch solche Aktionen können ein Indikator für erhöhtes Interesse durch den Nutzer sein. Diese Aktionen wurden im Logfile belassen. Sie kommen jedoch eher selten vor. In gerade einmal 6,25 % der Sessions in der Stichprobe konnte ein Klick auf die Aktionsleisten erkannt werden. Tab. 5.1 stellt die absoluten Häufigkeiten und das Verhältnis zwischen Transaktionen und Sessions dar.

Aktion	Transaktionen	Sessions	Verhältnis
Cite	63	40	1,58
Export	70	51	1,37
Mail	29	22	1,32
Bookmark	54	17	3,18

Tab. 5.1: Verwendung der Aktionsleisten: Ajax-Aktionen

Am häufigsten wurde auf *Export* geklickt. *Bookmark* hingegen wurde besonders oft innerhalb derselben Session wiederholt. Die Nutzer meldeten sich hierfür dann an und erstellten ihre Merklisten. Oft entstand bei der interpretativen Analyse jedoch eher der Eindruck, die Nutzer probierten einfach aus, was bei einem Klick auf die Aktionsleisten passiert. Vielleicht waren sie auch auf der Suche nach dem Zugang zu der gefundenen Publikation. Manchmal werden in derselben Session unterschiedliche Ajax-Aktionen vollzogen. Die Nutzer klickten sich dann durch die Leiste durch, wahrscheinlich um zu ermitteln, was die Funktionen dahinter sind.

Kapitel 6

Suchtaktiken

6.1 Definitionen von Suchtaktiken

Für die intellektuelle Analyse der von den Nutzern bei der Suche verwendeten Suchtaktiken wurde zunächst auf Listen aus der Literatur zurückgegriffen. Basierend auf den „search tactics“ von Marcia Bates (Bates 1990, 579), hat Alastair Smith eine Reihe von „internet search tactics“ aufgestellt, die für diese Analyse genutzt wurden. Im Folgenden wird dargelegt wie die Logfiledaten interpretiert und den Taktiken nach Smith zugeordnet wurden.

Im Vorfeld muss jedoch noch kritisch angemerkt werden, dass in der Liste von Smith solche Suchtaktiken nicht aufgenommen wurden, die er als trivial oder nicht gute Praxis einordnet. Darunter zählt er zum Beispiel die Suchanfrage als vollständige Frage zu formulieren (Smith 2012, 11). In dieser Untersuchung ging es aber nicht darum, nur nach den guten Taktiken der Nutzer zu fragen, sondern vielmehr zu erkennen, wo Verbesserungspotential für EconBiz liegt. Daher sind eigentlich gerade solche Taktiken interessant, die Probleme verursachen. Die Liste nach Smith 2012 wurde also um einige weitere Taktiken ergänzt, die sich ebenfalls im Logfile finden ließen.

Für die Interpretation der Vorgänge im Logfile und die Abbildung dieser Vorgänge auf angewandte Suchtaktiken, werden im Folgenden die Kriterien dargelegt. Der Großteil der Suchtaktiken von Smith kann im Logfile gar nicht nachvollzogen werden, weil dazu eine direkte Beobachtung des Nutzers notwendig wäre. Dazu gehören im Wesentlichen Evaluierungstaktiken oder solche, wo die Intention im Vordergrund steht wie etwa „VALUEADD“ (Smith 2012, 10). Die Kategorisierung stützt sich auf die in den Logfiles verzeichneten eingegebenen Suchbegriffe und das Bewegungs- bzw. Klickverhalten, sowie darauf folgende und vorhergehende Eingaben.

Die ausgewählten Suchtaktiken lassen sich in verschiedenen Kategorien einteilen:

6.1.1 Einstiegstaktiken

Unter Einstiegstaktiken werden die verschiedenen Wege gezählt, auf denen die Nutzer in eine EconBiz-Suche einsteigen. Der Begriff „Taktik“ ist an dieser Stelle eigentlich falsch gewählt,

zumindest wenn man unter „Taktik“ bewusst angewendetes Verhalten zählt. Im Falle von *Teleport* kann davon nicht unbedingt ausgegangen werden. Zu den Einstiegstaktiken zählen folgende:

Teleport: (nach Smith 2012 : 10) meint das Hineinspringen in EconBiz aus einer Websuchmaschine heraus. Ankunft des Nutzers erfolgt hier entweder auf einem Einzeltreffer oder einer Suchergebnisseite, die in einer anderen Suchmaschine indexiert war.

URL: (nach Smith 2012 : 10) meint das Eintippen einer URL in den Browser. Diese Taktik wurde nur dann vermutet, wenn der Referer keine Suchmaschine war und auf der Seite des Referers kein Link nach EconBiz gefunden werden konnte.

Provider: (nach Smith 2012 : 10) ist bei Smith das Aufsuchen einer Webseite, die wahrscheinlich die gewünschten Informationen erhält. Übertragen auf den Kontext von EconBiz wurde hier darunter verstanden, dass der Nutzer speziell EconBiz für die Suche auswählt, daher gilt diese Taktik immer dann, wenn die Session auf der Homepage von EconBiz startete. Dabei kann der Nutzer sowohl von einer Suchmaschine, als auch über Multiplikatoren auf EconBiz gekommen sein. In weiterem Sinne zählt auch die Nutzung des Suchschlitzes auf der Homepage der ZBW zu dieser Taktik, da auch hier angenommen werden kann, dass der Nutzer mit dem Angebot einigermaßen vertraut ist, oder zumindest davon ausgeht, dass er hier auf eine Quelle für nützliche Informationen zu seinem Bedürfnisse gelangt ist.

Eine weitere Möglichkeit, wie Nutzer in EconBiz ankommen, sind Verlinkungen. So etwas kommt z.B. vor auf Wikipedia-Seiten oder auf der Webseite der Zeitschrift „Wirtschaftsdienst“, von der aus man z.B. direkt weitere Veröffentlichungen eines Autors suchen lassen kann.

Eine sechste, nur selten genutzte Möglichkeit ist die Suche nach bestimmten Themen durch einen Klick auf einen Link aus dem Standard Thesaurus Wirtschaft heraus zu beginnen.

6.1.2 Reformulierungstaktiken

Zu den Reformulierungstaktiken zählen Taktiken, die die Veränderung der Suchanfrage beschreiben. Sie können nur bei durch den Nutzer zumindest in gewissem Maße „selbstformulierten“ Anfragen angewendet werden, als nur bei Suchanfragen, die über den Suchschlitz gestellt wurden.

- Exhaust:** (nach Smith 2012 : 10) Das Hinzukommen weiterer Aspekte des gesuchten Themas, wie etwa die Einschränkung auf ein Land. Bei Known-Item-Suche etwa das Hinzunehmen eines bibliographischen Datums zu einem anderen.
- Reduce:** (nach Smith 2012 : 10) Das Gegenteil von Exhaust, also das Wegfallen von einem oder mehreren Aspekten bei der Reformulierung einer Suchanfrage (zum Beispiel die Einschränkung auf ein bestimmtes Land wieder zurücknehmen).
- Parallel:** (nach Smith 2012 : 10) Gleichzeitiges Suche nach Synonymen oder in verschiedenen Sprachen. Dies wurde ausgeweitet auf mehrere Suchanfragen, also auch das Wechseln von einem Ausdruck zum anderen fällt unter diese Taktik.
- Relate:** (nach Smith 2012 : 11) Der Wechsel zu einem verwandten oder ähnlichem Begriff. Die Trennung war etwas unscharf zu *Parallel*. *Relate* wurden verwendet, wenn der Nutzer einen anderen Aspekt des Themas wählte, nach dem er suchte. Beispielsweise: Ein Nutzer sucht erst nach „Russland“, dann nach „Wirtschaftssanktionen“, dann nach „Ukrainekrise“ etc. Er deckt damit verschiedene Aspekte seines Themas nacheinander ab. Wurde dagegen zu einer Anfrage ein weiterer Aspekt angehängt (in diesem Beispiel dann „Russland Wirtschaftssanktionen“) würde dies unter *Exhaust* fallen.
- Sub:** (nach Smith 2012 : 11) Der Wechsel zu einem Unterbegriff. Kann nur bei thematischen Suchen verwendet werden.
- Super:** (nach Smith 2012 : 11) Der Wechsel zu einem Oberbegriff. Kann nur bei thematischen Suchen verwendet werden.
- Block:** (nach Smith 2012 : 11) Der Versuch Begriffe aus der Suche auszuschließen, etwa die Verwendung von NOT-Operatoren (-) . Die Beurteilung ist in der Praxis schwierig, da ein Minuszeichen eventuell auch unabsichtlich eingefügt worden sein könnte.
- Correct:** (nach Smith 2012 : 11) Smith unterscheidet *Correct* als Überwachungstaktik, und *Spellcorrect*. Diese Unterscheidung wurde hier nicht gemacht. Die Taktik kann also sowohl durch Verbesserung der Suchanfrage, als auch durch Korrektur falsch angewendeter Suchfunktionalitäten auftreten (etwa das Zurücksetzen einer falschen Feldbeschränkung).
- Respace:** (nach Smith 2012 : 11) Das Ausprobieren verschiedener Punktierungs- oder Leerzeichen Variationen in der Suchanfrage. Etwas das Aufspalten eines Suchbegriffs in seine Einzelteile

Respell: (nach Smith 2012 : 11) Das Ausprobieren alternativer Schreibweisen eines Suchbegriffes. Die Unterscheidung zu *Correct* liegt hier darin, dass es evtl. mehrere Transkriptionsmöglichkeiten eines Namens gibt. Es ist also keine reine Rechtschreibkorrektur gemeint.

Phrase: (nach Smith 2012 : 11) Die Nutzung von Zeichen in der Suchanfrage, zur Phrasierung.

Truncate: Das Trunkieren von Suchbegriffen

Form Acronym: (nach Huang/Efthimiadis 2009, 80) Das Bilden einer Abkürzung oder das Suchen mittels einer Abkürzung eines Begriffs.

Extend Acronym: (nach Huang/Efthimiadis 2009, 80) Das Ausschreiben eines zuvor abgekürzten Begriffes.

Substring: In diesem Kontext das Verkürzen einer Suchanfrage indem einfach Begriffe gelöscht wurden. Die Unterscheidung zu *Reduce* liegt darin, dass kein Aspekt des Suchthemas fortgenommen wird, sondern nur die Genauigkeit der Formulierung. Diese Taktik wurde exklusiv für Known-Item-Suchanfragen verwendet, in denen ein zuvor vollständige Titelangabe verkürzt wurde, um mehr Treffer zu erlangen.

Superstring: Das Gegenteil von *Substring*, exklusiv bei Known-Item-Suchanfragen, bei denen eine zuvor unvollständige Titelangabe präziser gemacht wird.

Questing asking: Formulierung der Suchanfrage als natürlichsprachige Frage.

Reorder Words: Verändern der Reihenfolge der Suchbegriffe in der Suchanfrage.

Stoppwort als Relation: Das Hinzunehmen von Stoppwörtern, um eine semantische Beziehung zwischen zwei Suchbegriffen herzustellen auf natürlichsprachigem Weg.

Zu den Taktiken, die zusätzlich zu Smiths aufgenommen wurden, zählen *Form/Extend Acronym*, *Stoppwort als Relation*, *Substring* und *Superstring*. Das Nutzen von Abkürzungen ist jedoch keine sehr oft verwendete Taktik. „Stoppwort als Relation“ beschreibt Fälle, in denen die Nutzer auf natürlichsprachigem Wege versuchen, der Suchmaschine aufzuzeigen, was für eine semantische Verbindung die beiden Suchwörtern wünschenswerter Weise in den Ergebnissen haben sollten. Auch diese Taktik ist jedoch eher selten aufgetreten.

Substring und *Superstring* gehen zurück auf eine Untersuchung von (Huang u. Efthimiadis 2009, 80), die sich an einer automatischen Ermittlung von Reformulierungsstrategien versucht haben. Bei ihnen ist es die Unterscheidung, ob zwischen einer ersten und einer zweiten

Suchanfrage etwas zum gleichen String dazugekommen ist (*Superstring*), der alte String aber noch enthalten, oder ob aus dem ersten String Teile entfernt wurden (*Substring*), um nur noch mit den Teilen des ersten Strings als Suchanfrage weiterzusuchen. Diese Unterscheidung ist im Rahmen der intellektuellen Interpretation unterkomplex, daher wurde die Definition zwar übernommen, jedoch nur auf bestimmte Fälle übertragen. Sie dienen als Beschreibung für die Taktik bei einer Known-Item-Suche die Ausführlichkeit der Titeleingabe zu ändern. Das Wechseln der Stichwörter hin zu einem anderen bibliographischen Datum, etwa das Hinzunehmen des Autornachnamens zur Suchanfrage, fällt dahingegen unter *Exhaust* bzw. *Reduce*. Das komplette Wechseln von einem bibliographischen Datum zum anderen (bei der ersten Suchanfrage der Autornachname, bei der zweiten nur Titelstichwörter) fiel dagegen unter *Relate*.

6.1.3 Suchfunktionstaktiken

Zu den Suchfunktionstaktiken zählen solche Taktiken, die im Wesentlichen beschreiben, welche Suchfunktion der Nutzer bei der Transaktion gewählt hat. Auch hier sind viele Suchtaktiken hinzugekommen, da sie sich speziell mit den Suchfunktionalitäten von EconBiz befassen und daher nicht auf Literatur gestützt sind. Hierzu zählen sowohl die in dieser Untersuchung als „Chaining“ bezeichneten Taktiken, bei denen durch Klick auf einen Feldeintrag eine Suche nach diesem Feldeintrag ausgelöst wurde, als auch der Wechsel von diesen Klickaktionen auf eine erste „selbstformulierte“ Anfrage im Suchschlitz von EconBiz. Dazu gehört auch die Nutzung der Facettenfilter oder der Wechsel der Suchfeldeinschränkung. Das Umsortieren der Ergebnisliste nach anderen Aspekten ist ebenso eine Taktik aus diesem Feld. Zunächst hier die Chaining-Taktiken, benannt in Anlehnung an das Informationssuchverhalten „Chaining“ nach Ellis, bei dem im Schneeballsystem ausgehend von einer gefundenen Publikation oder einem Literaturverzeichnis weitergesucht wird (Ellis 1989, 176):

Chain-Autor: Das Suchen nach weiteren Ergebnissen des gleichen Autors, durch Klick auf den Link auf dem Autornamen eines Treffers.

Chain-Series: Das Suchen nach weiteren Ergebnissen aus derselben übergeordneten Publikation durch Klick auf den Link der übergeordneten Publikation eines Treffers.

Chain-Subject: Das Suchen nach weiteren Treffern mit dem gleichen Schlagwort durch Klick auf den Link auf einem Schlagwort eines Treffers.

Chain-Type: Das Suchen nach weiteren Treffern der gleichen Publikationsform durch Klick auf den Link auf der Publikationsform-Information eines Treffers.

Chain-Institution: Das Suchen nach weiteren Treffern mit der gleichen verantwortlichen Institution durch Klick auf die Institutionsinformation eines Treffers.

Chain-similarAuthor: Das Klicken auf einen der rechts auf einer Einzeltreffer-Seite in EconBiz angebotenen ähnlichen Treffern, unter der Rubrik „Ähnliches Autorenspektrum“.

Chain-similarSubject: Das Klicken auf einen der rechts auf der Einzeltreffer-Seite in EconBiz angebotenen ähnlichen Treffern unter der Rubrik „Ähnliches Themenspektrum“.

Chain-Dublette: Das Klicken auf eine Dublette des bereits gefundenen Treffers auf einer Einzeltreffer-Seite rechts unter einer der beiden angebotenen Rubriken mit ähnlichen Treffern.

Daneben gibt es weitere Taktiken zur Beschreibung der verwendeten Suchfunktionalitäten:

Filter anwenden: Das Hinzufügen eines Filters auf eine Suchergebnisseite aus den Facettenfiltern auf der linken Seite in EconBiz.

Filter entfernen: Das Entfernen eines Filters auf einer Suchergebnisseite aus den Facettenfiltern auf der linken Seite. Die Unterscheidung von dieser Taktik vom bloßen Zurückgehen ist nicht immer eindeutig, wenn nur ein Filter gesetzt wurde oder der zuletzt gesetzte Filter entfernt wurde. War dies jedoch kurz vor dem Abbruch der Suche in EconBiz insgesamt, wurde meist als Zurückgehen gewertet.

Umsortieren der Ergebnisse: Das Umstellen der Trefferreihenfolge auf einer Suchergebnisseite. Diese Taktik ließ sich leider nicht in allen Fällen nachweisen, da hierzu keine neue URL nachgeladen wurde, solange der Nutzer nur die erste Suchergebnisseite aufruft. Ein technisch bedingter Reload war also in vielen Fällen nicht von einem Umsortieren der Ergebnisse zu unterscheiden. Erst wenn eine weitere Suchergebnisseite aufgerufen wird, kann man eine Veränderung in der URL erkennen.

Suchschlitz nutzen: Das Abschicken einer ersten initialen Suchanfrage über den Suchschlitz von EconBiz. Diese Taktik wurde auch vergeben, wenn erkennbar nach einem neuen Thema (oder einem anderen Known Item) gesucht wurde.

Erweiterte Suche nutzen: Das Abschicken einer initialen Suchanfrage über die Suchfunktion der erweiterten Suche, auch das bloße Auswählen und betrachten dieser Suchfunktionalität.

Field: Das Ändern der Suchfeldeinschränkung.

6.2 Bisheriger Forschungsstand zu Suchtaktiken

Die Änderungen von Suchanfragen wurden bereits in einigen Studien thematisiert. Viele Studien beschäftigten sich dabei mit der Findung von Taktiken und Strategien und nur einige mit den Veränderungen im Verlauf eines Suchprozesses. Häufig wurden hierfür Logfile-Analysen verwendet (Xie u. Joo 2010, 2191). Xie/Joo beobachteten und interviewten eine Reihe von Nutzern bei ihren Informationssuchen und fanden heraus, dass die Nutzer am meisten Energie in das Evaluieren von Ergebnissen und die Erlangung des Zugangs zu diesen Ergebnissen investieren, als etwa in die Formulierung von Suchanfragen (Xie u. Joo 2010, 2200). Spink et al. 2000 untersuchten unter anderem die Reformulierung von Nutzern der Websuchmaschine Excite und beobachteten, dass die meisten Nutzer ihre Suchanfragen nicht oft reformulierten (Spink u. a. 2000, 326). In einer späteren Studie zeigen die Ergebnisse jedoch häufigere Reformulierungen an (Jansen u. a. 2007b, 333). Sie stellten zudem fest, dass sich das Spezifizieren und Generalisieren die Waage hielt mit einer leichten Tendenz zu Spezifizierungen (Jansen u. a. 2007b, 334). Das komplette Wechseln der Suchbegriffe für ein Informationsbedürfnis wurde hier per Definition ausgeschlossen. Auch Boldi et al. 2011 verwendeten einen automatischen Classifier für die Beobachtung von Suchanfragereformulierungen und konnten so beobachten, dass der Wechsel der Suchbegriffe am häufigsten vorkam, und dass Spezifizierungen der Suchanfragen häufiger sind, als Generalisierungen (Boldi u. a. 2011, 15). Das Verhalten scheint dennoch abhängig vom System (oder der Kollektion) zu sein. Denn Petras et al. 2014 untersuchten das Reformulierungsverhalten in der European Library und stellten dort eine entgegengerichtete Reformulierungshäufigkeit fest, hier waren Generalisierungen der Suchanfragen am häufigsten (Petras u. a. 2014, 5). Huang/Efthimiadis untersuchten mit einem automatischen Classifier die Reformulierungen in einer Suchmaschine und die Klickmustern, die diesen folgten. Sie stellten fest, dass die Korrektur von Rechtschreibfehlern, das Ausschreiben von Akronymen und das Hinzufügen von Termen zur Suchanfrage häufiger zu einer Trefferauswahl führten, wenn die erste Suchanfrage keine Trefferauswahl zur Folge hatte. Wurde bereits bei der ersten Suchanfrage ein Treffer ausgewählt, so ist der Wechsel eines Suchterms, das Umstellen der Wörter in der Suchanfrage oder das Hinzufügen von Worten am häufigsten einer weiteren Trefferauswahl vorausgegangen (Huang u. Efthimiadis 2009, 83).

Die Nutzung der facettierten Filterfunktionen in verschiedenen Katalogen wurde bereits einige Male untersucht. Die Forschung zeigt, dass Facettenfilter eher selten und als Ergänzung zur Stichwortsuche genutzt werden (Niu 2013, 173). Die Nutzung ist auch hier abhängig von der Implementierung des Features. So stellte Niu beim Vergleich zweier unterschiedlicher Implementierungen in Bibliothekskatalogen unterschiedliche Quoten der Nutzung fest (Niu

2012, 71f.), was sich dadurch erklärte, dass die Nutzer die Facetten in einem Fall nur zur Verfeinerung einer Suche nutzen konnten, während im anderen Fall damit auch Browsing möglich war. Facettenfilter wurden in dieser Untersuchung häufiger für komplexere und offene Suchen verwendet (Niu 2012, 138).

Studien, die die Nutzung der Verlinkungen in OPACs zwischen den einzelnen Detailansichten untersuchen, was in etwa vergleichbar mit dem Chaining wäre, konnten nicht gefunden werden.

6.3 Ergebnisse zu Suchtaktiken

6.3.1 Reformulierungstaktik

Zu den Reformulierungstaktiken gehören alle, bei denen eine selbstformulierte oder auch automatisch generierte Anfrage vom Nutzer modifiziert wird. In 295 Sessions wurde mindestens eine Reformulierungstaktiken angewandt. Das entspricht einer Quote von 16,16 % in der Stichprobe. In 41,96 % der Sessions, in denen der Suchschlitz genutzt wurde, wurde mindestens eine Reformulierungstaktik verwendet. Es wurde also in nicht mal der Hälfte der Sessions, die die Suche nutzten, auch reformuliert.

Insgesamt wurden 641 Mal Reformulierungstaktik angewendet, das entspricht einer Quote von 41,9 % aller über den Suchschlitz gestellten Suchanfragen. Wenn Reformulierungen genutzt wurden, dann im Mittel gleich 2,1 Mal in der Session. Die am häufigsten angewendete Reformulierungstaktik war die Hinzunahme eines Aspekts/Begriffs zur Suchanfrage (*Exhaust*, dicht gefolgt vom Wechsel eines Begriffs in der Suchanfrage (*Relate*). Die Hinzunahme eines Aspekts kam auch am häufigsten in verschiedenen Sessions vor.

Der Unterschied zwischen dem Vorkommen in verschiedenen Sessions und dem Vorkommen insgesamt ist bei *Relate* und *Exhaust* am größten. Diese Taktiken wurden am häufigsten mehrmals innerhalb derselben Session verwendet.

6.3.1.1 Verwendung unterschiedlicher Reformulierungstaktiken

Wenn mehrmals in einer Session reformuliert wurde, wurden dafür oft verschiedenen Taktiken angewendet. Im Schnitt waren es 1,7 unterschiedliche Strategien. Wiederholt wurde eine Taktik im Schnitt 0,44 Mal pro Session, wobei der Median hier bei 0 liegt, d.h. dass in der Hälfte der Fälle die Taktik gar nicht wiederholt wurde in einer Session. Die Nutzer verwenden also oft unterschiedliche Taktiken innerhalb der Sessions.

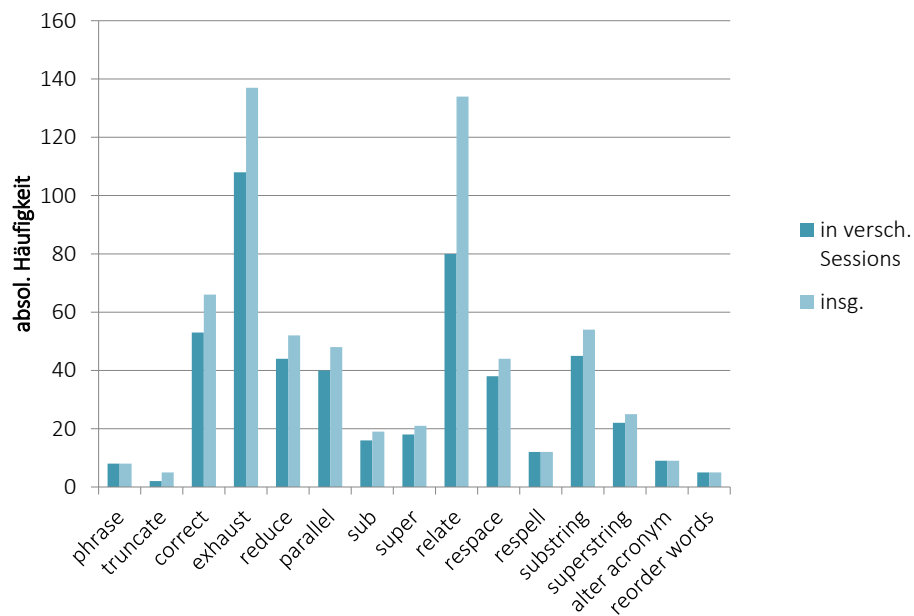


Abb. 6.1: Absolute Häufigkeiten der Reformulierungstaktiken

6.3.1.2 Reduzieren, Erweitern oder Wechseln?

Betrachtet man die Suchanfrage-Reformulierungen, so zeigt sich, dass das Erweitern der Suchanfrage (durch Hinzunahme eines Aspekts, etc.) sich mit dem Wechsel der Suchworte (*Parallel*, *Relate*) die Waage hält. Die Reduzierung einer Suchanfrage (das Wegnehmen von

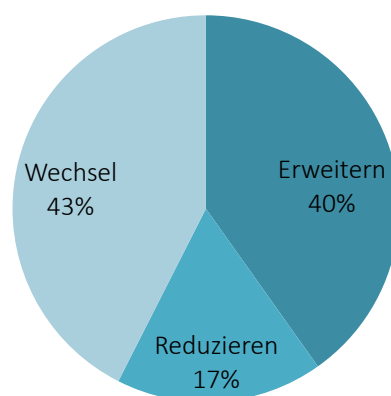


Abb. 6.2: Anteile der groben Reformulierungstaktiken

Suchbegriffen oder das Verkürzen der Anfrage) kommt dagegen deutlich seltener vor. Das Reformulierungsverhalten unterscheidet sich jedoch auch hinsichtlich des Suchtyps (s. 7.3.3). In thematischen Suchen wird viel häufiger der Suchbegriff gewechselt (54 %), der Anteil der Erweiterungen der Suchanfrage ist mit 38 % ähnlich, wie in der Gesamtheit aller Reformulierungen. Die Reduzierung der Anfrage hingegen sinkt auf 8 % der Reformulierungsvorgänge ab (p-Wert: 0,00, alpha: 0,05).

Dagegen wird in Known-Item-Suchen in fast einem Drittel der Fälle die Suchanfrage auf irgendeine Weise verkürzt und nur in einem Viertel der Fälle wird der Suchbegriff gewechselt. Bei Known-Item Suchen kommt dies aber auch dadurch zustande, dass der Wechsel des Suchbegriffes mit dem Wechsel des Suchfeldes einhergeht: wurde zuerst mit der Titelangabe gesucht, muss zum Autor gewechselt werden, um als eine „Wechsel“-Taktik zu gelten, das Hinzunehmen des Autors zum Titel wäre dagegen eine Erweiterung. Das Erweitern der Suchanfrage (44 %) überwiegt in Known-Item-Suchen das Reduzieren (30 %).

6.3.1.3 Sucherfolg der Reformulierungstaktiken bei thematischen Suchen

Erfolgsmessung stellte sich in der Logfile-Analyse insgesamt etwas schwierig heraus. Da durch die Struktur von EconBiz nicht alle Klickdaten ermittelt werden können, kann im Falle eines fehlenden Klicks auf einen Treffer nicht von einem Misserfolg ausgegangen werden (s. 4.4.3). Bei thematischen Suchen ließ sich der Erfolg also nur bemessen, wenn ein Klick auf einen Einzeltreffer erkennbar war und der Misserfolg nur im Falle von Nulltreffer-Ergebnissen sicher festzustellen.

Für viele Reformulierungstaktiken ist die Fallzahl so gering, dass kaum aussagekräftige Erfolgsquoten berechnet werden könnten. Bei den häufigsten Taktiken jedoch eher.

Die höchste Quote mit anschließender (sichtbarer) Trefferauswahl ergab sich bei *Exhaust*. In gut einem Drittel der Fälle wurde hier anschließend bei einer thematischen Suche ein Treffer angeklickt. Am zweitbesten schnitt *Parallel* ab mit gut einem Viertel Klicks. Bei *Relate* und *Reduce* liegt die Quote bei 17 %.

In Anbetracht dessen, dass die Nutzer bei thematischen Suchen häufiger den Begriff wechseln, als einen hinzufügen, muss die hohe Klickrate bei *Exhaust* unterstrichen werden. Die Erweiterung der Suchanfrage führte bei thematischen Suchen sicherer zu einem Klick, als das Wechseln des Suchbegriffs.

Die Nulltrefferquoten unterscheiden sich kaum voneinander für alle vier Taktiken liegt sie zwischen 13-16 %. Es kann also nicht davon gesprochen werden, dass eine Taktik besonders oft zum Misserfolg führte.

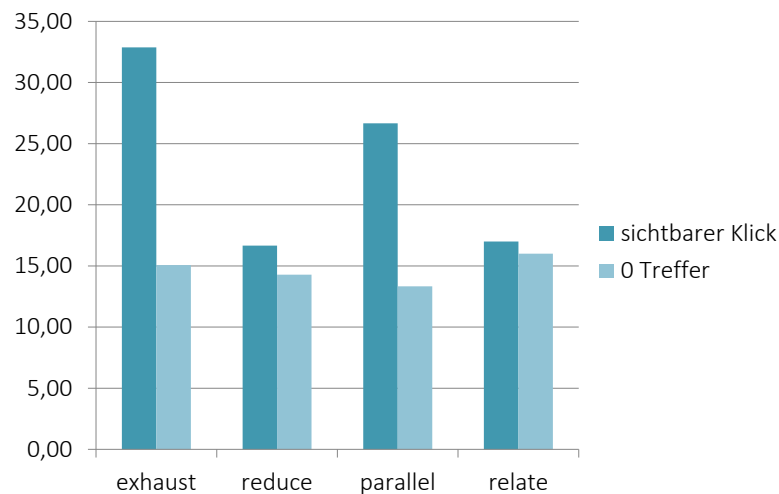


Abb. 6.3: Quoten der Erfolgs- und Misserfolgswfälle für die häufigsten Reformulierungstaktiken

6.3.2 Chaining-Taktiken

Chaining wurde in 1.008 verschiedenen Sessions angewendet, das entspricht 55,23 % der Stichprobe. Insgesamt wurde 1.341 Mal eine Chaining-Taktik verwendet. In mehr als der Hälfte der Sessions wurde also mindestens einmal irgendwo „weitergeklickt“. Im Durchschnitt wurde pro Session 1,3 Mal eine Chaining-Taktik eingesetzt. Am häufigsten wurde das Klicken auf die übergeordnete Publikation genutzt. Diese macht allein 47,05 % der verwendeten Chaining-Taktiken aus und kam in 57,34 % der Sessions mit mindestens einer Chaining-Taktik vor. Die zweithäufigste Chaining-Taktik war der Klick auf einen Autor. Dies geschah aber nicht einmal mehr halb so oft, wie der Klick auf die übergeordnete Publikation. Nur in 22,02 % der Sessions mit Chaining-Taktiken wurde der Autor geklickt, insgesamt 18,49 % der Chaining-Taktiken entfielen auf den Klick auf den Autornamen.

6.3.2.1 Sucherfolg bei Chaining-Taktiken

Auch hier lässt sich Erfolg wieder nur schlecht am Klickverhalten festmachen. Die Raten, bei denen die Nutzer auf einer durch Chaining entstandenen Suchergebnisseite einen Treffer auswählten, liegen zwischen 10,6 % bei der übergeordneten Publikation und 12,96 % bei der Institution. Beim Autor wurde in 11,87 % sichtbar etwas ausgewählt. Am schlechtesten ist die Rate wenig überraschend beim Chaining der Publikationsform mit 0,89 % sichtbarer Trefferauswahl. Hier wurde nur in einem Fall je ein Treffer ausgewählt, was stark darauf hindeutet, dass dem Klick auf die Publikationsform meist ein Missverständnis vorausging.

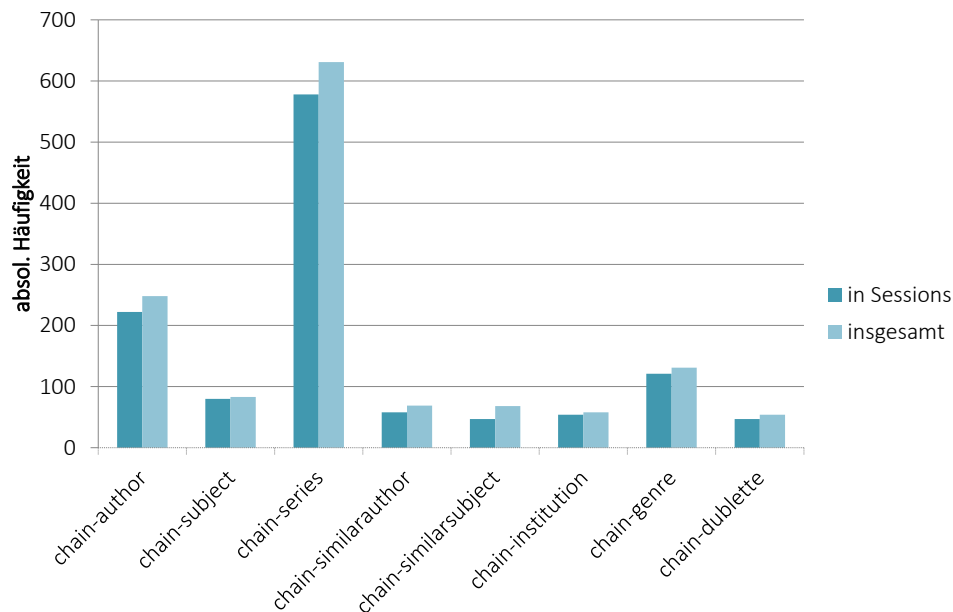


Abb. 6.4: Absolute Häufigkeiten der Chaining-Taktiken

6.4 Chaining vs. Suchschlitzsuche

Schon während der interpretativen Analyse entstand schnell der Eindruck, dass die Nutzer in EconBiz stärker im Chaining-Verfahren suchen, als mit selbstformulierten Suchanfragen. Nachdem dieser Eindruck entstanden war, wurde über die Gesamtheit der Sessions gezählt, in wie vielen Sessions eine Suchergebnisseite aufgerufen wurde, in der durch den URL-Zusatz „limit=20“ eine Chaining-Aktion stark angenommen werden kann. In 342.556 Sessions wurde eine Suchergebnisseite aufgerufen ohne diesen Zusatz. Dagegen wurde in 355.657 Sessions mindestens einmal eine Suchergebnisseite aufgerufen, mit diesem Zusatz. Das Verhältnis scheint damit beinahe ausgeglichen, sofern hier eine genau Trennung ermöglicht wurde.

In der Stichprobe lässt sich das Verhältnis besser beobachten, wenn man betrachtet welche Aktionen zu den Suchergebnisseiten geführt haben (s. Abb. 6.5).

Es wurden 3.828 Aktionen ausgeführt um (erstmalig) auf eine Suchergebnisseite zu kommen oder diese zu verändern. Auf der Ebene einzelner Aktionen unabhängig von Sessions entstanden 41,07 % der Suchergebnisseiten durch das Abschicken einer Suchanfrage, 30,38 % durch Chaining, 14,97 % durch das Auswählen einer indexierten Suchergebnisseite in einer Suchmaschine, und weitere 12,51 % durch Filteraktionen auf einer Suchergebnisseite. Das Verändern der Feldbeschränkung wurde nur in wenigen Fällen (41) vorgenommen. Hier sind

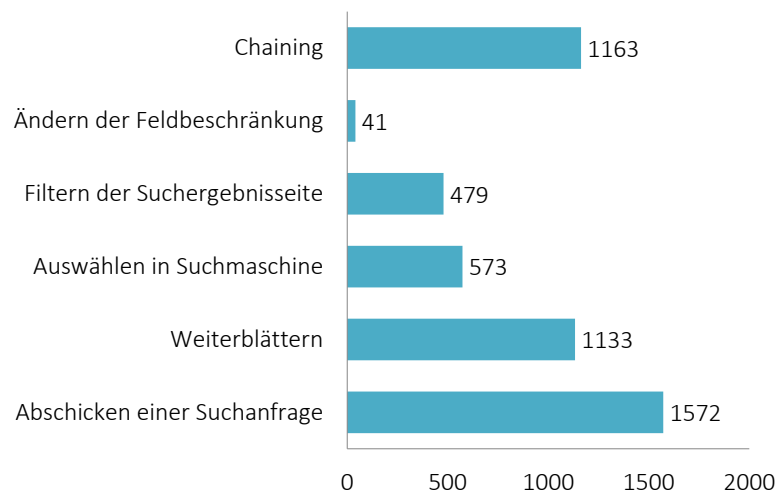


Abb. 6.5: Absolute Häufigkeiten von Aktionen bei Suchergebnisseiten

auch Fälle enthalten, wo die Feldeinschränkung nach Chaining durch den Nutzer zurückgesetzt wurde. Eine aktive Nutzung einer Feldbegrenzung für die Suche war also in nicht einmal 1 % der Suchergebnisse der Fall. In 1.133 Fällen wurde weitergeblättert auf einer der durch die oben aufgeführten Aktionen entstandenen Seiten, das entspricht einer Quote von 29,6 %, also nicht einmal ein Drittel der Suchergebnisseiten.

Das Verhältnis sieht anders aus, betrachtet man die Daten auf Sessionebene: In 1.519 Sessions wurde entweder eine Suchanfrage abgeschickt oder eine Suche durch Klicken ausgelöst. In 8,77 % (160) Sessions wurde sowohl eine Suchanfrage abgeschickt, als auch durch Klicken gesucht. Dagegen wurde nur in 38,52 % (703) Sessions mindestens einmal eine Suchanfrage abgeschickt, aber in 53,97 % (985) der Sessions wurde mindestens einmal durch Chaining eine Suchergebnisseite aufgerufen. In 29,75 % der Sessions wurde also nur durch Abschicken einer Suchanfrage gesucht, dagegen in 45,21 % nur durch Chaining.

Wurde also der Suchschlitz für die Suche genutzt, geschah das intensiver, als bei dem Suchen durch Durchklicken, auch wenn weniger Nutzer den Suchschlitz verwenden, als sich einfach nur Durchklicken. Die Kombination kommt dabei eher selten vor.

6.5 Transaktionsorientiertes Chaining?

Lässt man die Einstiegstaktiken und das bloße Nutzen des Suchschlitzes (also initiale Suchanfragen) heraus, so ist die Chaining-Taktik durch den Klick auf die übergeordnete Publikation

die am häufigsten in unterschiedlichen Sessions verwendete Taktik. In vielen Sessions entstand bei der interpretativen Analyse der Eindruck, dass der Klick auf die übergeordnete Publikation auch ein Versuch sein könnte, herauszufinden, wie man an die bereits gefundene Publikation heran kommt.

Nur in 108 von 578 Fällen war ein Klick auf einer durch den Klick auf die übergeordnete Publikation entstandenen Suchergebnisseite erkennbar. Es fehlen hier allerdings die Zahlen für Suchergebnisseiten, die erst nach Weiterblättern einen Klick bekamen, ursprünglich jedoch durch einen Klick auf den Link bei der übergeordneten Publikation entstanden sind.

In einer Reihe von Sessions fiel auf, dass die Nutzer nach dem Klick auf die übergeordnete Publikation überdurchschnittlich viele Suchergebnisseiten durchblättern, um dann letztlich wieder genau die Publikation auszuwählen, von der sie ursprünglich kamen. Zwar ist es denkbar, dass sie zwischendurch andere angeklickt haben, was sie von EconBiz wegführte und so nicht im Logfile auftaucht, jedoch warum klicken sie dann bei der bereits gefundenen wieder auf den Titel und nicht die Verfügbarkeitsprüfung? Das erscheint wenig plausibel, weshalb man annehmen kann, dass auch zwischendurch kaum Klicks erfolgt sein werden.

Noch einen Hinweis darauf, dass oft die Suche nach dem Zugang hinter dem Klick auf die übergeordnete Publikation steckt, ist das Verhalten, das sich zeigt, wenn die Nutzer zunächst nicht auf die übergeordnete Publikation, sondern auf die Publikationsform (z.B. „Artikel in Zeitschrift“) klicken. Sie erhalten dann eine Suchergebnisseite mit allen in einer Zeitschrift veröffentlichten Artikeln, merken anscheinend oft, dass dies nicht das ist, was sie erwartet haben, kehren dann zum Einzeltreffer zurück und klicken dann auf die übergeordnete Publikation. Hier zeigt sich also wahrscheinlich auch ein missverständlich formuliertes Label auf der Einzeltrefferseite. Die Nutzer scheinen hier also zu verstehen, beim Klick auf „Artikel in Zeitschrift/Article in Journal“ gelangen sie auf den Artikel oder die Zeitschrift des jeweiligen Treffers, den sie gerade vor sich haben.

Weitere Hinweise hätten Daten zur Online-Verfügbarkeit des gefundenen Einzeltreffers geben können, die leider im Rahmen dieser Untersuchung nicht erhoben wurden. Gerade bei nicht online verfügbaren Treffern wäre es denkbar, dass die Nutzer Informationen darüber gesucht haben, wo sie die Zeitschrift, in der der Artikel erschienen ist, finden können.

In dem Labeling-Test von EconBiz wurde zwar auch dieses Label getestet. Die Nutzer wurden jedoch nur gefragt, was das Datum bedeutet (die Schriftenreihe, in der der Artikel erschienen ist), nicht was sie erwarten, was passiert, wenn man dort hinklickt. Vor dem Hintergrund, dass ein gutes Fünftel der Anfragen an EconDesk aus dem Bedürfnis entstehen, den Zugang zu finden, erscheint es sinnvoll die These aufzustellen, dass die Taktik *Chain-Series* häufiger als transaktionsorientiert zu verstehen ist, als als informationsorientierte Handlung.

6.6 Filternutzung

Im Folgenden soll die Filternutzung, wie sie in der Stichprobe vorkam dargestellt werden. Die Filter wurden nur für die Suchergebnisseiten gezählt, bei denen bei der manuellen Analyse als Strategie *Filter setzen* vergeben wurde. In 450 Transaktionen von 212 Sessions wurde in der Stichprobe gefiltert. Das heißt in immerhin einem guten Zehntel (11,62 %) der in der Stichprobe untersuchten Sessions wurde vom Nutzer aktiv ein Filter eingesetzt.

Der Großteil der Filteraktionen wurde bei thematischen Suchen ausgeführt (74,37 %), nur ein Viertel bei Known-Item-Suchen. Direkt nach Filteraktionen wurde nur in 19,79 % ein Klick auf ein Treffer gemessen, das entspricht in etwa der allgemeinen Quote, in der Klicks im Logfile sichtbar waren.

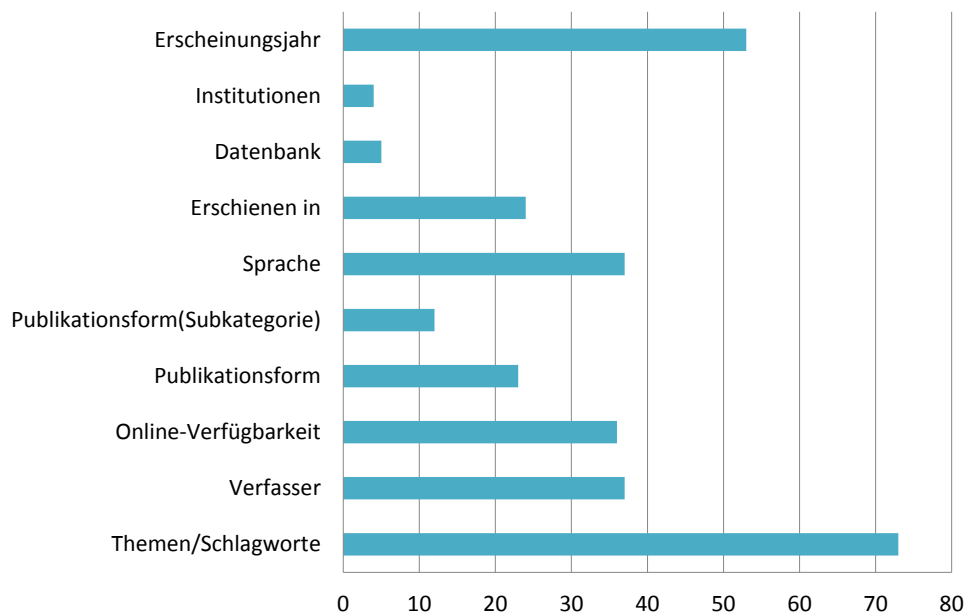


Abb. 6.6: Häufigkeiten der Filtertypen bei ersten aktiven Filterungen

6.6.1 Stichprobe

In den 450 Transaktionen, in denen aktiv ein Filter gesetzt wurde, kamen 617 Filter vor. Diese Zählung enthält also Doppelungen durch die Fälle, in denen auf einer bereits gefilterten Suchergebnisseite erneut gefiltert wurde. Am häufigsten war der Filter „Themen/Schlagworte“ gesetzt, gefolgt vom Filter „Sprache“ und „Erscheinungsjahr“ (kurz dahinter „Online-Verfüg-

barkeit“). Zählt man die beiden Filter zur Publikationsform zusammen, stehen sie bereits an dritter Stelle mit zusammen 81 Fällen.

Ein genaueres Bild ergibt sich, wenn man die Zahlen betrachtet, für solche Suchergebnisseiten, in denen nur ein Filter gesetzt wurde (s. Abb. 6.6), da in diesen Fällen eindeutig genau dieser Filter vom Nutzer genommen wurde, wohingegen bei der oberen Variante nicht klar ist, welcher Filter bereits da war und welcher neu hinzukommt.

Wenn eine Suchergebnisseite zum ersten Mal gefiltert wurde, dann am häufigsten mit „Thema/Schlagwort“. Am seltensten wurde nach Institutionen im ersten Schritt gefiltert und nach der Datenbank. Tabelle 6.1 zeigt die Häufigkeiten und Anteile der Filtertypen beim ersten Schritt der Filterung in der Stichprobe.

Filtertyp	Absolut	% von 617
Themen/Schlagworte	73	24 %
Erscheinungsjahr	53	17 %
Verfasser	37	12 %
Sprache	37	12 %
Online-Verfügbarkeit	36	11,7 %
Erschienen in	24	7,8 %
Publikationsform	23	7,6 %
Publikationsform(Subkategorie)	12	3,9 %
Datenbank	5	1,6 %
Institutionen	4	1,3 %

Tab. 6.1: Filtertypen beim 1. Schritt der Filterung

6.6.2 Annäherung über die Gesamtdaten

Um die Nutzung der Filterfunktion auf der Basis der gesamten Daten auszuwerten und dabei viele nicht aktiv durch Nutzer gesetzte Filter herauszunehmen, müssen folgende Bedingungen erfüllt sein:

1. Die URL sollte in der Session nur einmal gezählt werden, damit Aktionen wie das bloße Zurückgehen auf die vorherige Seite herausgefiltert werden.
2. Die gefilterte Suchergebnisseite kann nicht in der ersten Transaktion einer Session auftreten, die Transaktionsnummer muss also größer 1 sein.
3. Die Seitennummer der Suchergebnisseite sollte 1 sein, damit Filter nicht doppelt gezählt werden, wenn der Nutzer weitergeblättert hat.

4. Der Referer der Transaktion sollte „EconBiz“ sein, damit keine Suchergebnisseiten mitgezählt werden, die direkt in einer Suchmaschine aufgerufen wurden.

In 50.116 Sessions wurden Filter nach den vier Kriterien gesetzt, auf 103.258 unterschiedlichen URLs.

Umgerechnet auf die Gesamtzahl der Sessions bedeutet dies, dass in 1,46 % der Sessions ein Filter wahrscheinlich aktiv genutzt wurde. Auf den Seiten, auf denen wahrscheinlich händisch Filter gesetzt wurden, waren im Durchschnitt 1,44 Filter und es gab durchschnittlich 5,4 Filter pro Session. Beachtet werden muss jedoch, dass der zuerst gesetzte Filter ein zweites Mal gezählt wurde, wenn ein weiterer Filter hinzugesetzt wurde. Bei 5,4 Filter wurde also im Schnitt zweimal bis dreimal pro Session gefiltert. Drei Filtervorgänge zählt sechs Filter pro Sessions, wenn nur Filter hinzugefügt wurden.

Am häufigsten wurde der Filter „Thema/Schlagwort“ gesetzt, gefolgt vom Sprachfilter. Betrachtet man die Differenz zwischen absoluten Filterzahlen nach Typ und Filterzahlen des gleichen Typs auf denselben Suchergebnisseiten, zeigt sich, dass der Filter „Thema/Schlagwort“ am häufigsten mehrmals auf dieselbe Suchergebnisseite gesetzt wurde (also eine Suchergebnisseite wird zweimal nach Thema/Schlagwort gefiltert). Aus den Beobachtungen während der interpretativen Analyse kann dies damit erklärt werden, dass die Nutzer häufig sowohl den deutschen als auch den englischen Begriff als Filter verwenden.

6.7 Fazit zu Suchtaktiken

Die Untersuchung der Suchtaktiken zeigte, dass die Nutzer eher selten Reformulierungstaktiken einsetzen. Wenn sie dies tun, dann am häufigsten in dem sie einen weiteren Aspekt zur Suchanfrage hinzufügen. In thematischen Suchen nehmen sie dagegen häufiger einen anderen Suchbegriff.

Damit decken sich die Ergebnisse auch mit der bisherigen Forschung, die zeigte, dass die wenigsten Suchanfragen reformuliert werden. Mit der Quote von 41 % Reformulierungen unter den Suchanfragen, liegt die Quote auch bei EconBiz höher, aber noch ähnlich wie bei Jansen et al. 2007b bei 36,66 % (Jansen u. a. 2007b, 333).

Es zeigte sich auch, dass die Reduzierung der Ergebnismenge durch Hinzufügen weiterer Suchbegriffe häufig das Ziel der Nutzer ist. Auch die Klickquote nach einer Reduzierung der Ergebnismenge (durch *Exhaust*) war am höchsten unter den Reformulierungstaktiken. Die recht hohe Erfolgsquote bei *Exhaust* deckt sich mit den Ergebnissen von Huang u. Efthimiadis (2009), bei denen ebenfalls dem Hinzufügen von Suchbegriffen zur Suchanfrage häufig einen Klick auf der Suchergebnisseite folgte.

Die geringe Quote der Reformulierungen in den verschiedenen Sessions kommt jedoch auch dadurch zustande, dass die Nutzer in den meisten Sessions gar nicht auf den Suchschlitz als Funktion zurückgreifen, sondern sich eher durch EconBiz hindurch klicken, ohne eigene Suchanfragen zu formulieren. Nur in 39 % der Sessions in der Stichprobe wurde mindestens einmal den Suchschlitz genutzt, d.h. der Anteil bei allen Sessions ist noch geringer, wegen des hohen Anteils an Bounce-Sessions (s. 5.3).

Es wurde keine weitere Untersuchung der Nutzung von Chaining-Taktiken gefunden, da diese Funktionen auch EconBiz spezifisch sind. Häufig scheinen die Chaining-Taktiken jedoch der Versuch der Nutzer zu sein, den Zugang zu den gefundenen Publikationen zu erlangen (s. 6.5). Wie Xie/Joo beobachteten, macht dies den Großteil der Taktiken aus, die Nutzer in einem Suchprozess verwenden (Xie u. Joo 2010, 2200).

Durch die Analyse entstand der Verdacht, dass die Nutzer andere Erwartungen an die Suchfunktionalität haben, wenn sie auf die übergeordnete Publikation oder die Publikationsform („Artikel in Zeitschrift“) klicken. Dies sollte in einem weiteren Usability-Test genauer untersucht werden. Der Labeling-Test von 2012 fragte lediglich nach dem Verständnis des Datums hinter dem Label, nicht was die Erwartung der Nutzer war, wenn sie auf einen Link auf der Einzeltrefferseite klicken.

Die Nutzung der Facettenfilter war eher gering ausgeprägt. Auch hier ist der Aufruf einer Suchergebnisseite notwendig, was die Häufigkeit, in der sie genutzt werden schon deutlich reduziert. Wie auch bei Niu 2013 zeigte sich, dass Filter häufiger bei thematischen Suchen verwendet werden, als bei Known-Item-Suchen. Dies erklärt auch die Beliebtheit der Facette „Thema/Schlagwort“.

Sucherfolg bei den verschiedenen Taktiken stellte sich unterschiedlich häufig ein. Die Quoten eines Klicks oder Sucherfolgs nach Chaining-Taktiken sind deutlich geringer, als nach Suchanfragen über den Suchschlitz. Auch das Filtern ist noch erfolgreicher als Chaining. Das erklärt sich auch dadurch, dass Chaining meist bereits dann erfolgt, wenn eine Suche in Google bereits erfolgreich war. Da die Vermutung nahe liegt, dass viele Chaining-Einsätze transaktionsorientiert sind, unterscheidet sich diese Suche von den anderen. Es wird genau das gesucht, was als Einzeltreffer aus Google aufgerufen wurde. Um wieder dorthin zu gelangen, können die Nutzer auch einfach im Browserverlauf zurückgehen. Dies wurde jedoch nicht als Erfolg gewertet.

Kapitel 7

Known-Item-Suche

Das Konzept der Known-Item-Suche wird in vielen Untersuchungen von Information Retrieval Systemen, insbesondere auch bei der Untersuchung von OPACs verwendet. Die Definitionen variieren je nach Kontext etwas. Für diese Untersuchung wurde folgende Definition gewählt, aufgrund derer die Kriterien für eine Einteilung der Suchen in EconBiz vorgenommen wurden. Eine Known-Item-Suche ist eine Suche, bei der:

1. der Suchende ein bestimmtes Werk (bzw. eine Kopie davon) finden möchte.
2. der Suchende nicht unbedingt genau wissen muss, dass das gesuchte Werk existiert.
3. der Suchende nicht alle bibliographischen Daten zu dem gesuchten Werk kennen muss (Lee u. a. 2006, 12).
4. der Suchende jedoch mindestens mit Teilen von bibliographischen Daten als Ausgangspunkt sucht.

7.1 Erkennen einer Known-Item-Suche

Die Definition muss letztlich operational sein, da die eigentliche Intention und der Kontext des Nutzers in einer Logfile-Analyse nicht beobachtet werden können. D.h. üblicherweise würde man davon ausgehen, dass Suchanfragen, die bibliographische Daten enthalten, wie Autornamen, Titelstichwörter, Jahreszahlen oder Ähnliches, als Known-Item-Suchen zu verstehen sind (so etwa bei (Wildemuth u. O'Neill 1995, 265)). Einzelne Begriffe können zwar Bestandteil von Titeln sein, werden aber in dieser Untersuchung als thematische Suche eingestuft, sofern sich nicht anhand von etwaigen Reformulierungen mehr Hinweise auf ein Known-Item ergeben. Im Zweifelsfall wird also gegen eine Known-Item-Suche entschieden.

In der vorliegenden Untersuchung ergab sich mit dieser Herangehensweise jedoch recht frühzeitig ein Problem.

Die Untersuchung erfolgte auf Sessionebene. Dadurch ließen sich sowohl die ausgewählten Einzeltreffer, als auch die eingegebenen Suchanfragen und die Art der Eingabe erkennen.

Zentrales Problem für die Einordnung als Known-Item-Suchen war das Unwissen über die initiale Suchanfrage, da diese in den meisten Fällen in Google erfolgt war und der Nutzer über eine Einzeltreffer oder eine Suchergebnisseite in EconBiz einstieg. In vielen Fällen wurde im Anschluss daran in Econbiz nicht mehr weiter mit neuen, vom Nutzer selbst formulierten Anfragen gesucht, sondern weitergeklickt. Zur Illustration ein typisches Beispiel:

Der Nutzer hat etwas in Google gesucht (was ist nicht mehr nachvollziehbar). Er kommt auf eine Einzeltrefferseite in EconBiz. Nun klickt er auf den Autornamen der gefundenen Publikation. EconBiz sucht nun mit einer automatischen Anfrage alle Publikationen dieses Autors aus dem Index. War das eine Known-Item-Suche? Würde man nur die Suchanfrage betrachten, wäre der Fall einer Known-Item-Suche stark anzunehmen. Denn sie enthält lediglich Vor- und Nachname eines Autors, mit der Feldeinschränkung auf das Feld: Autor. Es sind jedoch mehrere Gründe denkbar, warum der Nutzer den Autornamen angeklickt hat:

- a) Er sucht eigentlich nach einer anderen Publikation dieses Autors, als die bereits über Google gefundene. Dann wäre es eine Known-Item-Suche.
- b) Er hat die bereits gefundene Publikation gesucht, will aber eigentlich mehr über das Thema wissen und hofft, dass der Autor noch mehr dazu veröffentlicht hat. Die Known-Item-Suche fand in der Suchmaschine statt. Jetzt erweitert sich das Bedürfnis auf thematische Information: also keine Known-Item-Suche mehr.
- c) Er hat ursprünglich thematisch gesucht, den Einzeltreffer gefunden und hofft unter dem Autor mehr zu finden. Es war also immer eine thematische Suche.
- d) Er erhofft sich irgendwelche weiteren Informationen wie er an die Publikation heran gelangen kann. Es geht ihm also um Informationen zum Zugang zu der gefundenen Publikation, was noch immer als Known-Item-Suche gelten kann.
- e) Er kennt EconBiz vielleicht nicht und probiert einfach mal aus, was passiert, wenn man da klickt.

Zu Beginn tendiert man zu b). Hinter einer Known-Item-Suche steckt in den meisten Fällen ein anderes Informationsbedürfnis, das größerer Natur ist, als die Suche nach einem bestimmten Dokument (Lee u. a. 2006, 13).

Die Ambivalenz kann auch durch Klickdaten kaum aufgehoben werden: Nach dem Klick auf den Autornamen...

1. wird eine andere Publikation des Autors ausgewählt.
2. wird auf die alte Publikation zurückgegangen.

3. wird abgebrochen.
4. wird die Suche reformuliert.

In Fall 1) kann der obige Fall a) eingetreten sein: Er hat eigentlich eine andere Publikation gesucht. Er könnte jedoch genauso gut b) sein, er sucht nach einem Thema und hat eine weitere Publikation dieses Autors gefunden, die ebenfalls passend ist.

In Fall 2) lässt sich ebenso wenig eine gute Entscheidung treffen, was der Fall gewesen sein könnte. Er könnte entweder nichts für das Thema passendes entdeckt haben oder nicht die Publikation entdeckt haben, nach der er eigentlich sucht.

In Fall 3) kann man ebenfalls nicht genau wissen, ob a) oder b) oder c). Bricht er die Suche auf der Ergebnisseite ab, scheint er nichts Passendes für irgendeinen der Fälle gefunden zu haben. Die Suche ist also in jedem Fall nicht erfolgreich. Um jedoch zu diesem Schluss zu kommen, wäre es notwendig alle Klicks nachverfolgen zu können. Es fehlen im Logfile jedoch die Informationen zu allen Klicks.

In Fall 4) kommt es auf die Art der Reformulierung der Suche an. Hier wäre die Einkategorisierung der zweiten Suchanfrage von Belang. Wählt er einen eher thematischen Einstieg ist es entweder b) oder c), wählt er einen auf bibliographischen Daten, würde ich a) annehmen. Durch diese Ambivalenz der Bewertung wurde die Betrachtung von Known-Item-Suchen auf durch den Nutzer selbstständig formulierte Anfragen über den Suchschlitz beschränkt. Es gilt also folgende operationale Definition für die Bewertung als Known-Item-Suche:

1. Die Suchanfrage ist selbst formuliert, nicht vom System konstruiert.
2. Die Suchanfrage enthält bibliographische Angaben, wie Titelstichwörter, Jahreszahlen, Namen, ISBN etc.

Zur Identifizierung der Known-Items wurden die Klickdaten und die Suchergebnisseiten bei der Reproduzierung der Suche, sowie die Reformulierungsvorgänge betrachtet. Es wurde in EconBiz und Google, sowie Google Scholar geprüft, ob ein solches Known-Item mit den Angaben der Suchanfragen vorhanden ist, um herauszufinden, ob die Suchanfragen zu einer Known-Item-Suche gehörten.

7.2 Aktueller Forschungsstand zu Known-Item-Suchen

Mit Known-Item-Suchen wird sich schon seit vielen Jahrzehnten in der Bibliotheks- und Informationswissenschaft beschäftigt. Die Ergebnisse schwanken und sind deutlich abhängig von den Definitionen und Operationalisierungen der Forschenden und der Informationssysteme,

die untersucht wurden. So reichen die Anteile von Known-Item-Suchen an den Suchen in den Suchsystemen und Katalogen von um die 20 % bis hin zu 79 % in Studien, die Logfile-Analyse nutzen (Wildemuth u. O'Neill 1995, 267) und zwischen 30 % - 67 % in Studien, die Nutzer befragten (Wildemuth u. O'Neill 1995, 269). Auch in neueren Untersuchungen zeigt sich, dass Known-Item-Suchen einen hohen Anteil an den Suchen in Discoverysystem oder Bibliothekskatalogen haben (s. z.B. (Krebs 2013, 93), (Chapman u. a. 2013), (Rulik 2014, 36)).

Das Äquivalent von Known-Item-Suchen in Websuchmaschinen sind navigationsorientierte Suchanfragen. Auch hier machen diese Art von Suchanfragen einen bedeutenden Anteil der Suchen in Websuchmaschinen aus (s. z.B. (Lewandowski 2011, 357)). Es kann also auch für diese Untersuchung in EconBiz angenommen werden, dass der Anteil an Known-Item Suchen groß ist, vielleicht sogar die Hälfte oder mehr der Suchvorgänge in EconBiz ausmacht.

Die Untersuchungen von Known-Item-Suchen sind häufig Logfile-Analysen (Wildemuth u. O'Neill 1995, 266). Meist erfolgt die Klassifizierung der Anfragen über die Inhalte der Suchanfragen allein oder die Feldbeschränkung, die für die Anfrage verwendet wurde (Wildemuth u. O'Neill 1995, 268). Die Veränderungen der Suchanfragen im Sessionverlauf hat Krebs (2013) bei seiner Untersuchung mit einbezogen, jedoch nicht genauer quantifiziert. Er untersuchte auch, welche bibliographischen Informationen in den Suchanfragen der Nutzer verwendet wurden und stellte fest, dass der Titel oder Titelstichwörter unter den Inhalten der Known-Item-Suchanfragen am häufigsten Vorkommen (Krebs 2013, 97). Gleiches gilt auch für die Untersuchung von Rulik ((Rulik 2014, 41)). Es zeigt sich auch hier, dass die Ergebnisse stark abhängig sind von den Operationalisierungen der Forschenden und den Systemen, die betrachtet wurden. Titel- und Autorensuchen oder Titel- und Autorenstichwörter sind jedoch auch in allen älteren Studien die Hauptbestandteile der Suchanfragen oder Suchbeschränkungen gewesen (Wildemuth u. O'Neill 1995, 268).

Erfolg einer Known-Item-Suche wird in Kataloguntersuchungen meist gemessen, indem Trefferzahlen verglichen werden oder Nulltreffer-Seiten gezählt. In ihrer Zusammenfassung der Untersuchungen zu Known-Item-Suchen kommen Wildemuth/O'Neill zu der Erkenntnis, dass in etwa einem Viertel der Known-Item-Suchen ein Misserfolg stattfindet (Wildemuth u. O'Neill 1995, 270). Krebs hatte sogar fast ein Drittel Nulltreffer-Ergebnisse (Krebs 2013, 93).

Rulik betrachtete auch die Retrievaleffektivität des Discoverysystems beluga. Etwa 74 % der Known Items wurden dabei auf den ersten Rang gerankt (Rulik 2014, 43). In einer vergleichenden Untersuchung der Retrievaleffektivität von Suchmaschinen bei navigationsorientierten Suchanfragen kam Lewandowski auf 84 % Erfolg bei Rang 1 bei Google (Lewandowski 2011, 359).

7.3 Ergebnisse

7.3.1 Anteil der Known-Item Suchen

Welchen Anteil hatten Known-Item-Suchen an den Suchen in EconBiz? Diese Frage zu beantworten ist nicht ganz so leicht, wie zunächst gedacht. Nur in 38,52 % (703) aller untersuchten Sessions der Stichprobe wurde der Suchschlitz aktiv genutzt. Fast zwei Drittel klickten sich also nur irgendwie durch EconBiz durch und es ist daher nicht möglich eine genauere Antwort zu finden (s. 7.1). In 20,33 % (371) der Sessions wurde der Suchschlitz für die Durchführung einer Known-Item Suche ausgeführt.

Ein besseren Überblick gibt die Betrachtung, wie oft eine initiale Suchanfrage über den Suchschlitz gestellt wurde und wie oft dies eine Known-Item-Suche war. Die Betrachtung der initialen Suchanfragen nimmt Verzerrungen durch reformulierte Anfragen heraus und gibt so einen besseren Überblick, als wenn man alle Suchanfragen für die Betrachtung mit einbeziehen würde. Es wurden 862 Mal über den Suchschlitz initiale Suchanfragen in den Sessions gestellt. Davon können 491 als Known-Item-Suchen betrachtet werden. Das entspricht also einem Anteil von 56,96 % der initialen, über den Suchschlitz gestellten Suchen. Von den 39 % die also überhaupt eine Suche ausführen, sind die meisten tatsächliche eine Known-Item-Suche.

Im Schnitt wurden 1,2 verschiedene Known-Item-Suchen in den Sessions ausgeführt, in denen Known-Item-Suchen über den Suchschlitz vorgenommen wurden. Es kommt also einige Male vor, dass in einer Session mehrere verschiedene Known-Items recherchiert werden.

Betrachtet man die Suchergebnisseiten, die über Suchmaschinen aufgerufen wurden, kann man vermuten, dass die Dunkelziffer an Known-Item-Suchen höher liegt, wie im Folgenden dargelegt wird.

7.3.1.1 In Suchmaschinen indexierte Suchergebnisseiten

Insgesamt 521 Suchergebnisseiten wurden direkt in einer anderen Suchmaschine ausgewählt. Dies geschah in 451 verschiedenen Sessions. 79,46 % der über eine Suchmaschine aufgerufenen Suchergebnisseiten waren innerhalb ihrer Session die erste aufgerufene Seite. Durchschnittlich hatten diese Seiten 2.609 Treffer, wobei eine stark linksschiefe Verteilung zu sehen ist. Der Median liegt bei 10 Treffern pro Seite, 18,81 % hatten sogar nur einen einzigen Treffer und zwei Drittel nur bis zu 25 Treffer.

Schaut man sich die „Suchanfragen“ an, die bei dieser Art von Seite angegeben sind, fällt auf, dass die Hälfte nur Namen (Vor- und Nachname) von Personen enthält. Ein weiteres Viertel enthält den Titel einer Publikation oder einer bestimmten Zeitschrift. Unter den 14 %

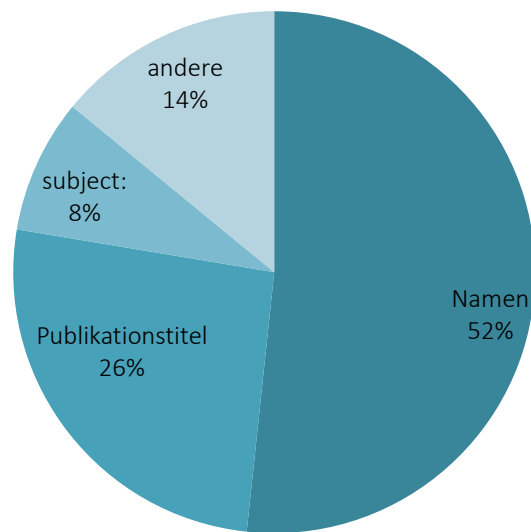


Abb. 7.1: Inhalte der Suchanfragen auf in Google indexierten Suchergebnisseiten

anderen Inhalten sind einzelne Stichwörter oder die Namen von Institutionen (s. Abb. 7.1). Der Verdacht liegt also nahe, dass hier in mehr als drei Viertel der Fälle in der Suchmaschine eine Known-Item-Suche ausgeführt wurde. Weitere Hinweise geben die voreingestellten Filter, die sich auf solchen Suchergebnisseiten finden.

Diese geringe Trefferzahl liegt nicht zuletzt daran, dass auf gut einem Viertel dieser Seiten (23,42 %) bereits Filter voreingestellt sind, in den meisten Fällen nur einer. Am häufigsten waren das die Filter „Verfasser“ (29,92 %) und „Themen/Schlagworte“ (28,34 %). Die Filter „Publikationsform(Subkategorie)“ und „Erschienen In“ liegen beide bei 8,66 %. Am seltensten war der Filter „Online-Verfügbarkeit“ (0,78 %) auf einer solchen Suchergebnisseite.

Geht man davon aus, dass die automatischen „Suchanfragen“ auf den von Suchmaschinen indexierten Seiten einen Hinweis geben können, was eventuell in der Suchmaschine gesucht wurde, so kann man vermuten, dass 78 % dieser Sessions als „transaktionsorientiert“ gelten können, da die Nutzer in EconBiz den Zugang zum gesuchten Item finden wollen würden.

7.3.2 Elemente der Suchanfrage bei Known-Item-Suchen

Um zu ermitteln, welche Elemente die Nutzer für eine Known-Item-Suche verwendeten, wurden alle über den Suchschlitz formulierten Suchanfragen und deren Reformulierungen betrachtet. 748 verschiedene Suchanfragen wurden als Known-Item-Suche gekennzeichnet. In einer Session mit Known-Item-Suche wurden im Schnitt 1,9 verschiedenen Suchanfragen für

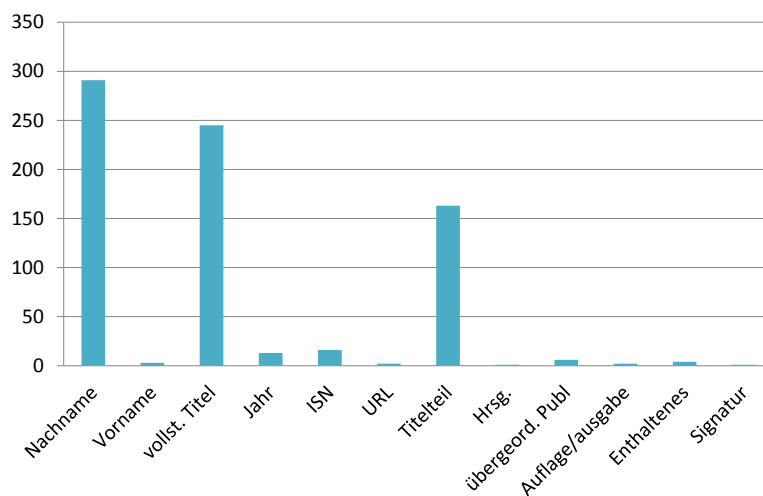


Abb. 7.2: Absolute Häufigkeiten der Elemente in Known-Item-Suchanfragen

eine Known-Item-Suche gestellt. Im Schnitt enthält eine Suchanfrage 1,5 verschiedene Elemente. Der Median liegt bei 1. Die Suchanfragen sind im Schnitt 4,72 Wörter lang, der Median liegt bei 3 Worten.

Am häufigsten trat der Nachname des Autors in verschiedenen Suchanfragen auf. Er wird in 38,48 % als Element für die Suchanfrage einer Known-Item-Suche verwendet. Mehr als zwei Drittel der Anfragen enthalten unter anderem den vollständigen Titel der gesuchten Publikation (36,75 %). Fast ein Drittel enthält nur Teile (einzelne Stichwörter oder den Anfang des Titels) in der Anfrage (31,03 %). Zusammen enthalten also 67,78 % der Anfragen ausschließlich oder unter anderem den Titel oder Teile davon. Der Vorname des Autors wird nur in 16,11 % der Fälle mit verwendet, in 2,93 % nur als Initiale.

Betrachtet man die Verteilung der Elemente für die initialen Suchanfragen (ohne Reformulierungen), steigt der Anteil des vollständigen Titels sogar auf 42,67 %, wohingegen nur noch 26,78 % nur Teile des Titels enthalten. Der Autornachname kommt in 37,24 % der initialen Suchanfragen vor, der Vornamenanteil bleibt bei 16,11 %. Betrachtet man nur die Verteilung der Elemente in Suchanfragen, die bereits eine Reformulierung sind, so enthalten die meisten hier den Autornachnamen (40,66 %), am zweithäufigsten enthalten sie Teile des Titels (38,46 %). Nur 26,37 % enthalten den vollständigen Titel.

Dies legt nahe, dass die Nutzer zunächst den vollständigen Titel eingegeben haben, auf diese Weise oft nicht den richtigen Treffer finden konnten und daher die Suchanfrage verkürzten.

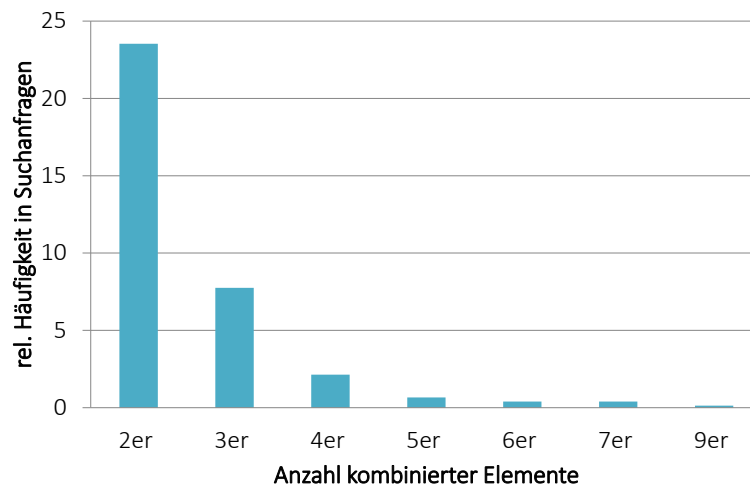


Abb. 7.3: Häufigkeit von Elementkombinationen

7.3.2.1 Alleinstehenden Elemente einer Known-Item-Suchanfrage

In 486 Suchanfragen wurde nur ein Element verwendet. Das entspricht fast zwei Dritteln der Suchanfragen bei Known-Item-Suchen (64,84 %). Am häufigsten bestanden die Suchanfragen mit nur einem Element aus der vollständigen Titelangabe (47,53 %). Es folgen in der Häufigkeit Suchanfragen mittels Titelbestandteilen zu einem Drittel (32,78 %). Der Nachname des Autors wurde bei Suchanfragen mit nur einem Element zu 12,55 % alleinstehend verwendet. Zusammengenommen decken diese drei Elementtypen bereits 93 % aller Suchanfragen ab, die nur ein einziges Element enthielten.

66,19 % der Suchanfragen aus Known-Item-Suchen, die nur ein Element enthielten, waren initiale Suchanfragen. Betrachtet man die Verteilung der Elemente nur über diese initialen Suchanfragen, so unterscheiden sich die Anteile der Elementtypen kaum von den Anteilen in allen Suchanfragen mit nur einem Known-Item Element. Der vollständige Titel ist in 51,71 %, Titelteile in 30,53 % und der Nachname des Autors in 12,77 % enthalten.

7.3.2.2 Mehrere Elemente in Known-Item-Suchanfragen

In 35,02 % (262) der Anfragen waren mehrere Elemente enthalten. Die meisten Kombinationen enthielten nur zwei verschiedene Elemente. Die höchste Anzahl an Elementen waren 9 in einer Anfrage. Hier wurde eine komplette Literaturangabe in den Suchschlitz hineinkopiert. Wenn mehrere Elemente in der Suchanfrage einer Known-Item-Suche waren, enthielt die Suchanfrage in 87,79 % der Fälle den Nachnamen eines Autors. Mehrere Nachnamen wurden

nur in 13,74 % der Fälle verwendet.

Vornamen kommen immerhin noch in 45,42 % der Suchanfragen mit mehreren Elementen vor. Der vollständige Titel kommt nur in 16,79 % (44) der Fälle vor. Titelteile hingegen in 28,62 % (75) der Fälle.

Die Jahresangabe der Veröffentlichung hat einen viel höheren Anteil unter Suchanfragen mit mehreren Elementen, sie liegt hier bei 23,28 % (61 Fälle), wohingegen sie gesehen über alle Anfragen nur in 8,25 % vorkommt und unter den Suchanfragen mit nur einem Element gar nicht.

7.3.2.3 Elementkombinationen in Known-Item-Suchanfragen

Die häufigste Kombination von Elementen ist, wenig überraschend, die Kombination von Vor- und Nachnamen eines Autors. 45,41 % (119) der Suchanfragen mit mehreren Elementen enthielten unter anderem den Vor- und Nachnamen eines Autors. Allein den Nachnamen und den Vornamen enthielten 30,53 % (80) der Suchanfragen mit mehreren Elementen. Das sind immerhin noch 10,69 % aller Known-Item-Suchanfragen. Der Vorname kommt mit nur einer Ausnahme immer in Kombination mit dem Nachnamen vor.

Das Element, das am häufigsten mit der Jahresangabe kombiniert wurde, ist der Nachname eines Autors. Beide Angaben finden sich in 16,41 % der Suchanfragen mit mehreren Elementen. Das zweithäufigste Element zusammen mit einer Jahresangabe ist ein Teil des Titels. Diese beiden Elemente finden sich in 6,49 % (17 Fälle) aller Suchanfragen mit mehreren Elementen. Der vollständige Titel ist nicht weit abgeschlagen mit 5,34 % (14 Fälle).

Der vollständige Titel wird am häufigsten kombiniert mit dem Nachnamen des Autors. Er kommt jedoch viel häufiger allein stehend vor. Einzelne Titelstichwörter werden, wenn sie kombiniert werden, ebenfalls am häufigsten mit dem Nachnamen des Autors kombiniert. Die Unterscheidung von Namen wäre also ein gutes Mittel Known-Item-Suchen von anderen Suchanfragen zu trennen.

7.3.2.4 Known-Item-Suchen ohne Titel, Autor und Jahresangaben

Wurde weder eine Namens-, noch eine Titel-, noch eine Jahresangabe in der Suchanfrage verwendet, so wurden eigentlich fast immer eine ISN-Nummern (15 Fälle) gesucht. In seltenen Fällen auch die übergeordnete Publikation (6) und im Falle einer Known-Item-Suche wurde nach etwas gesucht, von dem bekannt war, dass es in der Publikation enthalten ist. Der Fall war insofern interessant, als das erkennbar war, dass der Nutzer zwar wusste, dass es die Publikation gibt, und dass sie „Satellitenfotos von Europa“ enthält, aber kaum genaue Angaben zum Titel und Autor hatte.

7.3.3 Suchtaktiken bei Known-Item-Suchen

Für die Known-Item-Suchen wurden im Schnitt 1,66 verschiedenen Taktiken gewählt. Als Grundlage dienen hier nicht die verschiedenen Anfragen, sondern die verschiedenen Taktiken die über den Suchschlitz ausgeführt wurden. Es kann passieren, dass die gleiche Anfrage im Verlauf einer Session mit unterschiedlichen Taktiken erneut vorkommt. Die Taktiken wurden im Verlauf der Sessions vergeben. Wenn also z.B. ein Nutzer zunächst mit dem ganzen Titel sucht, dann einen Teil entfernt und dann doch wieder den ganzen Titel nimmt, so ist der Textstring der ersten Anfrage zwar identisch mit der dritten, jedoch nicht die Taktik.

Wie bereits unter 7.3.2 beobachtet, scheint es die Tendenz zu geben, dass die Nutzer vollständige Titelangaben im Laufe der Suche abkürzen zu Titelteilen. Betrachtet man die durchschnittliche Anzahl unterschiedlicher Elemente für initiale und reformulierte Anfragen so sieht man, dass der Durchschnitt sich wenig unterscheidet: 1,52 Elemente für initiale und 1,56 Elemente für reformulierte Anfragen.

Es konnten 399 Reformulierungsvorgänge bei Known-Item-Suchen identifiziert werden. Die am häufigsten verwendete Reformulierungstaktik bei als Known-Item-Suche gekennzeichneten Suchanfragen war *Exhaust* (35 % der Reformulierungen), also das Hinzunehmen eines weiteren Aspektes zur Suche. Diese Taktik bedeutet übertragen auf den Fall einer Known-Item-Suche:

Das Hinzunehmen einer weiteren, anderen Angabe aus den bibliographischen Daten des gesuchten Items. Hatte die erste Anfrage beispielsweise nur den Nachnamen enthalten, nimmt man dann noch eine Titelangabe mit hinzu.

Die zweithäufigste Reformulierungstaktik war hingegen *Substring*. Diese Taktik und ihr Gegenstück *Superstring* wurden speziell für den Fall von Known-Item-Suchen in die Liste der Strategien aufgenommen. *Substring* wurde in 16,54 % der Reformulierungsvorgänge angewendet.

Das Gegenstück *Superstring*, also das Erweitern der Titelangabe bis hin zur vollständigen Titelangabe, wurde in 8,27 % der Reformulierungsvorgänge vorgenommen. Weitere 10 % machen das Korrigieren von Fehlern bei der vorherigen Suchanfrage aus. Entweder die Nutzer bemerkten Tippfehler, oder sie bemerkten eine falsche Feldeinschränkung, die sie bei der zweiten Suchanfrage korrigierten. Das Herauslösen eines Aspektes aus der Anfrage (*Reduce*) macht nur noch 6,02 % der Reformulierungsvorgänge aus.

Dies zeigt, dass die Nutzer bei Reformulierungen von Known-Item-Suchen eher dazu neigen, eine andere bibliographische Angabe hinzuzunehmen, als sie in der ersten Suchanfrage genutzt hatten. Eine Alternative wäre die Taktik *Relate*, also das Wechseln des Feldes, z.B. statt nach dem Nachnamen nur nach dem Titel suchen. Diese Taktik wurde jedoch nur in

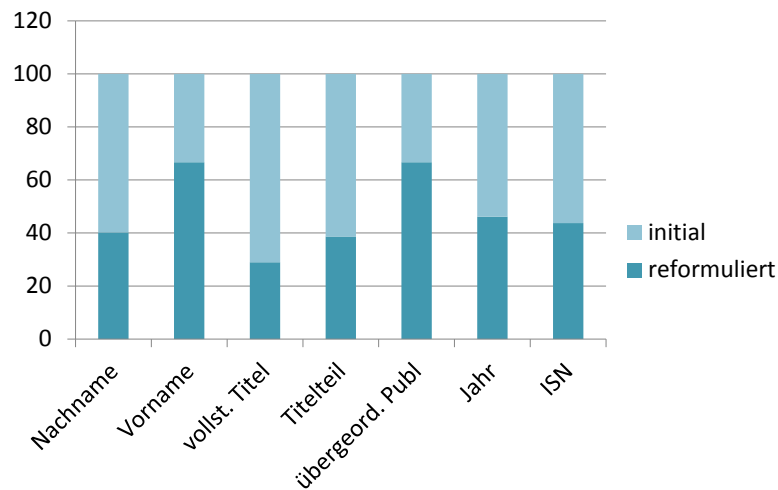


Abb. 7.4: Relative Anteile von reformulierten Anfragen bei Elementen von Known-Item-Suchanfragen

14 % der Fälle verwendet, also nicht einmal halb so oft wie *Exhaust*.

In reformulierten Suchanfragen war signifikant häufiger nur ein Teil des Titels enthalten und ebenso signifikant häufiger der Nachname des Autors in der Suchanfrage, als in den initialen Suchanfragen (p-Wert 0,003; alpha: 0,05). Der vollständige Titel war dagegen viel seltener in den reformulierten Suchanfragen, als in den initialen (25 % gegenüber 36 % in den initialen Suchanfragen) (s. Abb. 7.4). Auch dies deutet daraufhin, dass die Nutzer mit der vollständigen Titelangabe beginnen und bei Misserfolg die Angabe verkürzen.

7.4 Sucherfolg bei Known-Item-Suchen

Um die Erfolgsquote zu ermitteln wurden zunächst alle Transaktionen ermittelt, bei denen eine Known-Item-Suche gekennzeichnet wurde (s. Abb. 7.5).

Daraufhin wurden aus der Betrachtung solche Transaktionen ausgeschlossen, die Reload oder Zurückgeh-Aktionen waren. Es verblieben 1.017 Transaktionen in der Betrachtung. Fragt man jetzt über diese Transaktionen die Erfolgsquote ab, so wurde in 38,05 % der Transaktionen ein Erfolg gemessen.

Von allen Transaktionen, bei denen bei einer Known-Item-Suche eine Anfrage über den Suchschlitz abgeschickt wurde, wurde in 44,67 % ein Erfolg registriert. In Transaktionen, in denen ein Erfolg registriert wurde, wurde in 29,97 % der Fälle vom Nutzer das erste Ergebnis angeklickt und deshalb ein Erfolg vergeben, während in 47,54 % der Erfolgsfälle ein Erfolg vergeben wurde, weil das gesuchte Item oder ein Passendes auf dem ersten Rang stand. In

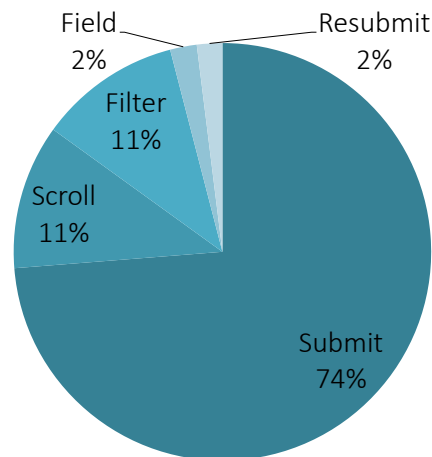


Abb. 7.5: Anteile der Aktionstypen bei Known-Item-Transaktionen

17 % der Erfolgsfälle wurde ein Ergebnis ab Rang 2 bis Rang 10 ausgewählt. In 3,36 % der Fälle wurde auf einer weiteren Suchergebnisseite etwas angeklickt.

Betrachtet man nur die initialen (ersten Anfragen) einer Suche, so liegt die Erfolgsquote bei immerhin 50,99 %. Gut die Hälfte der Known-Items wurde also bereits bei der ersten Anfrage auf den ersten Rang gerankt oder angeklickt.

Reformulierte Suchanfragen haben eine Erfolgsquote von 33 %, sind also weniger erfolgreich, als initiale Suchanfragen. Gründe hierfür könnten darin liegen, dass vielleicht etwas gesucht wurde, dass es gar nicht zu finden gab, die Nutzer jedoch nicht so leicht aufgegeben haben (s. 7.4.2).

In 68,02 % der Sessions mit einer Known-Item-Suche war zumindest eine Transaktion bei der Known-Item-Suche erfolgreich.

Die Erfolgsquoten sind erkennbar nicht sehr gut. In nicht einmal der Hälfte der Transaktionen wurde das gesuchte Item auf den ersten Rang gebracht oder vom Nutzer angeklickt. Im Folgenden wird das Ranking genauer betrachtet.

7.4.1 Ranking

Success@n wurde berechnet über alle Submit-Aktionen, beziehen also das Ranking der Known-Items auf der ersten Suchergebnisseite direkt nach einer über den Suchschlitz gestellten Suchanfrage mit ein (das sind 74 % der Known-Item-Transaktionen s.o.).

Bezieht man nur die Known-Items in die Bewertung ein, die auch gefunden wurden auf der ersten Suchergebnisseite in der Reproduzierung, so sieht die Erfolgsrate recht gut aus. Ein

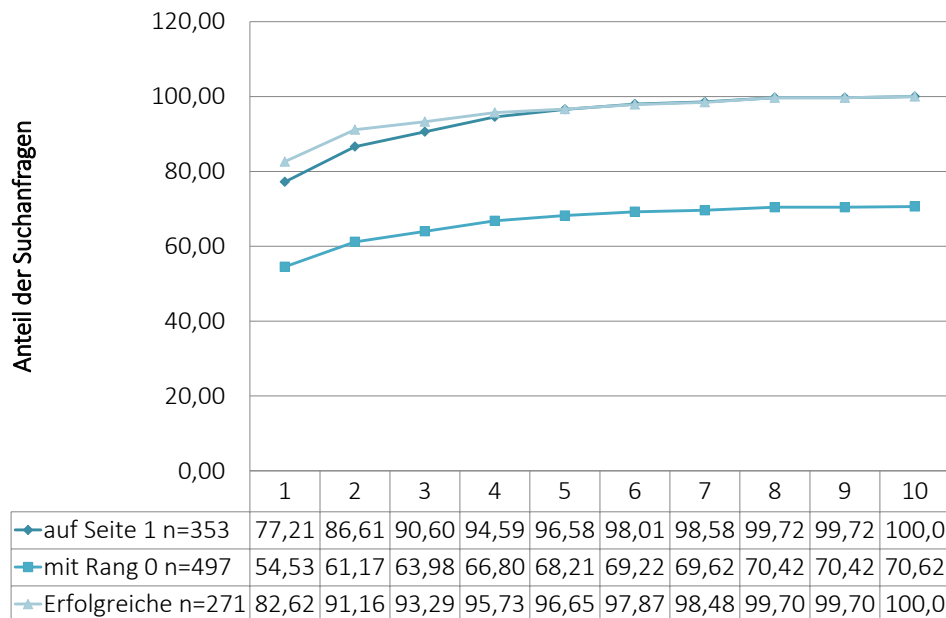


Abb. 7.6: Success@n für Known-Item-Suchanfragen

bisschen weniger als ein Viertel der Known-Items war nicht auf Rang 1. Die Kurven für erfolgreiche Suchanfragen (wo also das Known-Item angeklickt wurde oder Rang 1 gegeben war), unterscheidet sich kaum von der Kurve über die Known-Items auf Seite 1. Die Differenz entsteht, weil nicht alle Known-Items, die nicht auf Rang 1 standen auch angeklickt wurden und daher nicht als Erfolg gewertet wurden (s. 4.4.3).

Vollkommen anders stellt sich die Success@n-Rate bereits dar, bezieht man auch Known-Items mit ein, deren Rang nicht ermittelt werden konnte, obwohl sie in EconBiz auffindbar gewesen wären (Rang 0). Nur knapp die Hälfte der Known-Items wurde dann noch auf Rang 1 gefunden und fast ein Drittel wurde nicht auf der ersten Suchergebnisseite angezeigt. Der Mean Reciprocal Rank für alle Suchanfragen, deren Known-Item theoretisch auffindbar gewesen wäre, beträgt nur 0,59 und bestätigt damit die Ergebnisse des Success@n. Das ist zunächst kein gutes Abschneiden bei der Retrievaleffektivität bei Known-Item-Suchen in EconBiz.

Eine Erklärung für das schlechte Abschneiden wäre die Problematik der Datenerfassung. In einigen Fällen wurde der Klick des Nutzers als Rang notiert, obwohl das Known-Item auf Rang 1 stand. Die Dublettenproblematik beispielsweise führte häufig dazu, dass der vom Nutzer geklickte Einzeltreffer Rang 2 oder 3 hatte, obwohl der gleiche Treffer auch auf Rang 1 stand. Die hierdurch entstandenen Fehlerquote lässt sich im Nachhinein nicht mehr gut

abschätzen. Im Folgenden wird versucht zu ermitteln, was weitere Gründe für dieses doch durchwachsene Abschneiden gewesen sein könnten.

7.4.2 Gründe für den Misserfolg

Für alle Transaktionen, in denen ein Misserfolg registriert wurde, wurde in einem Drittel dieser Fälle (35,09 %) das Known-Item nicht auf der ersten Suchergebnisseite angezeigt, obwohl es in EconBiz theoretisch auffindbar gewesen wäre. 21,44 % der Suchergebnisseiten hatten 0 Treffer bei einer Known-Item-Suche. Das sind 46,53 % der nicht erfolgreichen Suchanfragen. In 6 % der Transaktionen mit Misserfolg, war das Item auf der erste Suchergebnisseite (> Rang 1), wurde aber nicht erkennbar angeklickt.

7.4.2.1 Unpräzise Suchanfragen

Wenn die Suchanfragen so unpräzise waren, dass das Known-Item auf Rang 1 kommen konnte, so müssten die Trefferzahlen bei nicht erfolgreichen Known-Item-Suchen häufig sehr hoch sein.

Die Anzahl gefundener Treffer durch Submit-Aktionen unterscheidet sich deutlich nach Suchtyp. Während bei thematischen Suchen im Mittel 252.733 Treffer erzielt wurden (Median 133), liegt die Trefferanzahl bei Known-Item-Suchen bei 5.821 Treffern (Median 5) (t-Wert: 0,0000036; alpha: 0.05). Known-Item-Suchen erzielten demnach deutlich weniger Ergebnisse, als thematische Suchen.

Der Durchschnitt bei erfolgreichen Submit-Aktionen von Known-Item-Suchen liegt bei 797 Treffern gegenüber 9.866 Treffern bei nicht erfolgreichen. Für die Trefferzahlen bei erfolgreichen und nicht erfolgreichen Known-Item-Suchen konnte jedoch kein signifikanter Unterschied in den Trefferzahlen festgestellt werden. Die Varianzen in der Stichprobe sind zu groß, um verlässliche Werte zu liefern. Es lässt sich also sagen, dass Known-Item-Suchen weniger Treffer erzielen, als thematische Suchen, jedoch nicht, dass ein Unterschied der Trefferzahlen beim Sucherfolg von Known-Item-Suchen zu sehen ist.

7.4.2.2 Fehlende Indexierung

In einem Drittel der Known-Item-Suchen wurde nach einem Item gesucht, das wahrscheinlich (nach eigener Prüfung) gar nicht in EconBiz auffindbar war, weil entweder die Angaben des Nutzers falsch waren oder das Item nicht indexiert ist. Dieser Umstand erklärt bereits 58,91 % der nicht erfolgreichen Transaktionen von Known-Item-Suchen, aber nicht das Abschneiden bei der Retrievaleffektivität, denn dort werden diese Transaktionen gar nicht betrachtet.

Durch das Hereinspringen aus anderen Suchmaschinen wissen viele Nutzer vielleicht nicht genau, was in EconBiz zu finden ist (s. 8.4).

7.4.2.3 Rechtschreibfehler

In allen Transaktionen, in denen eine Known-Item-Suche ausgeführt wurde, waren in 6,54 % der Fälle Rechtschreibfehler in der Suchanfrage. Dies ist also nur ein kleiner Anteil der Fehlerquellen für Misserfolge bei einer Known-Item-Suchen.

7.4.2.4 Falsche Feldbeschränkung

Um herauszufinden, wie oft eine fehlerhafte Feldbeschränkung die Ursache für einen Misserfolg der Known-Item-Suche war, wurde gezählt, wie häufig die Feldeinstellung nicht auf der Standardeinstellung stand, gleichzeitig ein Misserfolg und keine aktive Änderung des Nutzers der Feldbeschränkung vorlag. Diese Fälle machen zusammen 18,91 % der Misserfolge aus. Nimmt man als weiteres Kriterium Nulltreffer-Seiten hinzu sind sehr wahrscheinlich 12,36 % der Misserfolgswfälle durch eine falsche Feldbeschränkung zu erklären. Da sich in der interpretativen Analyse gezeigt hat, dass 0 Treffer jedoch nicht in jedem Fall einer irrtümlichen Feldbeschränkung auftreten, können auch die 18,91 % als gute Annäherung an die Zahl der Misserfolge durch eine falsche Feldeinstellung gelten. Betrachtet man dazu noch, welches Feld eingestellt war, so deckt sich die Verteilung ziemlich gut, mit der Chaining-Rate der einzelnen Felder. Am häufigsten war das Feld „Erschienen In“ eingestellt, halb so oft das Feld „Verfasser“. Auf Sessionebene ausgedrückt: in 12,74 % der Sessions mit Known-Item-Suche lag ein Misserfolg durch eine falsche Feldeinstellung vor, das entspricht etwa 3 % aller Sessions in der Stichprobe.

Misserfolg bei Known-Item-Suchen lässt sich also zum Großteil begründen durch eine unvollständige Sammlung. Rechtschreibfehler machten einen eher geringen Anteil aus, dagegen erklärt die falsche Verwendung der Feldbeschränkung einen guten weiteren Teil der Misserfolge.

Sofern nicht mehrere Misserfolgsgründe gleichzeitig aufgetreten sind, kommt man auf mindestens 15,64 % der nicht erfolgreichen Transaktionen, in denen die Retrievalleistung hinterfragt werden müsste. Abzüglich eventuell noch einiger sehr unpräziser Suchanfragen durch die Nutzer, die später eindeutig reformuliert wurden und daher zunächst einen Misserfolg auslösten.

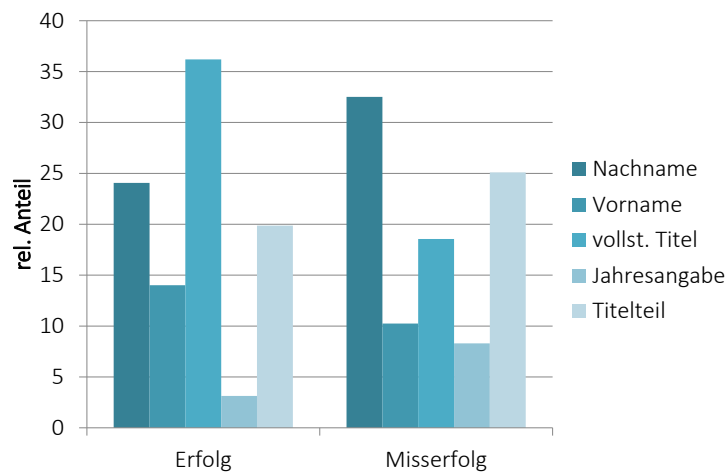


Abb. 7.7: Relativer Anteil der Elemente von Known-Item-Suchanfragen in erfolgreichen und nicht erfolgreichen Suchanfragen

7.4.2.5 Sucherfolg mit Suchanfrage-Elementen

Betrachtet man die in erfolgreichen und nicht erfolgreichen Suchanfragen verwendeten Elemente, so fällt auf, dass in den erfolgreichen Suchanfragen der Anteil der vollständigen Titelangaben signifikant höher ist, als in den nicht erfolgreichen (p -Wert: 0,00; α : 0,05). Hier ist wiederum der Anteil der Suchen mit einzelnen Titelstichwörtern signifikant höher. Auch die Nutzung des Nachnamens kommt häufiger in nicht erfolgreichen Suchanfragen vor, als in erfolgreichen. Die Differenz fällt jedoch nicht so deutlich aus (s. Abb. 7.7). Dies gibt also weitere Hinweise darauf, dass die Nutzung von Titelteilen weniger erfolgreich war, als den gesamten Titel zu verwenden.

Betrachtet man, in welchen Fällen der Misserfolg bei der Verwendung der Elemente durch fehlende Indexierung entstand, so zeigt sich, dass in etwa der Hälfte der Fälle, in denen die oben aufgeführten Elemente vorkamen jeweils eine fehlende Indexierung des gesuchten Known-Items den Misserfolg der Suche erklärt (s. Abb. 7.8).

7.4.3 Suche nach bereits gefundenen Treffern

Ein interessantes Phänomen ließ sich in einem guten Fünftel der gelungenen Known-Item-Suchen feststellen: die Nutzer suchten ein Known-Item, dass sie kurz zuvor bereits als Einzeltreffer aufgerufen hatten. Besonders häufig verwendeten sie hierfür die vollständige Titelangabe. Es liegt nahe anzunehmen, dass die Nutzer sogar den Titel aus dem Einzeltreffer in den Suchschlitz kopierten. Eine denkbar sinnvolle Erklärung für so ein Verhalten wäre die

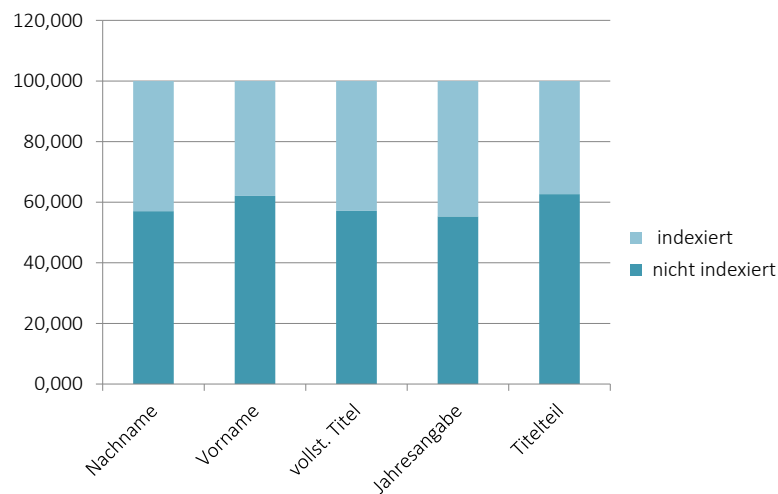


Abb. 7.8: Rel. Anteile der Misserfolgswfälle und Indexierungszustand in Transaktionen, pro Element

diffuse Hoffnung der Nutzer, weitere Informationen zum Zugang zu der bereits gefundenen Publikation zu erhalten, indem sie statt in Google in EconBiz danach suchen. Ein Abgleich der vorherigen Einzeltrefferaufrufe in einer Sessions mit den Suchanfragen kann also helfen, transaktionsorientierte von informationsorientierten (oder rein navigationsorientierten) Anfragen zu unterscheiden.

7.5 Fazit zu Known-Item-Suchen

Die Beobachtungen aus der Logfile-Analyse von EconBiz decken sich gut mit den bisherigen Untersuchungen zu Known-Item-Suchen in Online-Katalogen oder auch Discovery-Systemen. Sieht man von den vielen Fällen ab, in denen die Suche in EconBiz gar nicht „klassisch“ zum Einsatz kam, so fällt der Anteil der Known-Item-Suchen an den Suchanfragen mit 57 % doch ziemlich groß aus und bestätigt die Vermutung, dass ein Großteil der Suchen in EconBiz Known-Item-Suchen sind, vorausgesetzt, die Suche von EconBiz wird aktiv genutzt.

Auch die Beobachtung, dass der Titel in einem Großteil der Suchanfragen von Known-Item-Suchen enthalten ist, bestätigt Ergebnisse der anderen Studien (s. 7.2) auch für EconBiz. Die Nulltreffer-Quote liegt, verglichen mit Studien zu OPACs, mit gut einem Fünftel der Known-Item-Suchen niedriger als in anderen Studien, was zunächst positiv zu bewerten ist. Die Fehler durch Rechtschreibfehler in den Suchanfragen waren zumindest bei Known-Item-Suchen eher selten der Grund für einen Misserfolg. Das mag daran liegen, dass häufig der Titel

wahrscheinlich direkt in den Suchschlitz kopiert wurde. Sehr oft wurde einfach etwas gesucht, das in EconBiz nicht vorhanden ist. Hier wäre eine weitere Untersuchung lohnenswert, die ergründet, inwiefern die nicht indexierten Items eigentlich in die Sammlung von EconBiz gehören würden oder inwiefern die Nutzer sich schlicht an der falschen Adresse umgesehen haben (etwa bei ingenieurwissenschaftlichen Themen).

Die Retrievaleffektivität erreicht dagegen ziemlich schlechte Werte im Vergleich mit der Literatur. Auch beluga erreichte in der genannten Untersuchung 74 % auf dem ersten Platz. Selbst wenn man beachtet, dass in der Untersuchung die Kategorie der als „unsicher“ eingestuften Known-Item-Suchanfragen herausgelassen wurde, so würde deren Hinzunahme dennoch nicht auf die gleichen Werte wie in dieser Untersuchung fallen.

Abgesehen von den erst im Nachhinein aufgefallen methodischen Ungenauigkeit, dass Klickverhalten und Retrievaleffektivität im gleichen Datenfeld vermerkt werden mussten und daher Verfälschungen entstanden sind, liegt der Verdacht nahe, dass viele Misserfolge durch die falsche Feldeinschränkung entstanden sind. Diese Funktion gibt es weder bei der allgemeinen Websuche, noch in beluga. Angesichts der sehr geringen Nutzungshäufigkeit sollte ihre Notwendigkeit in Frage gestellt werden. Auch die sessionbasierte Betrachtung der Suchanfragen (und die damit auch rückwirkende Einkategorisierung als Known-Item-Suche von Anfragen, die nur anhand ihrer Reformulierungen als solche erkannt werden konnten), können das Ergebnis bei der Retrievaleffektivität in Frage stellen. Manche Suchanfragen waren vielleicht sehr unpräzise. Eine weitere Untersuchung sollte gezielt diese Frage in den Blick nehmen.

Ein Phänomen, das in keiner Studie erwähnt wird, ist das Wiedersuchen von bereits gefundenen Known-Items. Durch die besondere Konstellation, dass EconBiz eben auch Einzeltreffer von allgemeiner Websuche indexieren lässt und die Nutzer so mitten in die Suche hineingeworfen werden, ohne das angenommen werden kann, dass sie EconBiz oder die ZBW kennen, wird dies erklärt.

Das Zusammenspiel von allgemeiner Websuche und Online-Katalogen wurde in keiner der Studien zur Known-Item-Suche bisher thematisiert. Es zeigt sich aber, dass die Nutzung von EconBiz als reiner Content-Anbieter, der von außen angesteuert wird, die Nutzung als Literatursuchmaschine bisher deutlich überwiegt (s. 6.4). Der folgende Abschnitt beschäftigt sich mit Unterschieden im Suchverhalten abhängig vom Einstiegspunkt der Suche, und auch hier wird diese Nutzung von EconBiz deutlich.

Kapitel 8

Herkunft der Nutzer und Einstiegspunkte in EconBiz

8.1 Hypothesen und Forschungsstand zu Einstiegspunkten

Das Suchverhalten abhängig vom Ausgangspunkt der Suche zu betrachten, ist keine häufig eingenommene Perspektive bei der Untersuchung von Information Retrieval Systemen. Das mag auch daran liegen, dass die Möglichkeit, auf einem Einzeltreffer die Nutzung des Systems zu beginnen, meist gar nicht gegeben ist. Üblicher ist die Beobachtung eines unterschiedlichen Verhaltens eher im e-Commerce, wo Nutzerpfade daraufhin analysiert werden, wo sich Verkäufe verbessern lassen (Bucklin u. Sismeiro 2009, 41). Der sogenannten „Conversion-Funnel“ ist ein Modell für einen Such- und Kaufprozess (Abhishek u. a. 2012, 3). Er zeigt auch an, an welchen Stellen die Kunden vor dem Kauf den Prozess verlassen. Hierfür werden dann unter anderem auch „user paths“ oder „click streams“ analysiert. Die Ergebnisse sind meist jedoch auf einzelne Webseiten bezogen und lassen sich nur schlecht übertragen. Veröffentlicht werden verschiedene Modelle, wie sich Kundenverhalten messen lässt (Abhishek u. a. (2012), Wilson (2010)), die eigentlichen Untersuchungen werden jedoch meist nicht veröffentlicht.

Ortega/Aguillo untersuchten Unterschiede im Besucherverhalten einer Webseite (Aufenthaltsdauer und Klicks) in Abhängigkeit von der Herkunft der Nutzer. Sie beobachteten signifikant längere Sessions bei Nutzern, die von einer Suchmaschine kamen, als z.B. Nutzern, die auf der Startseite der Webseite begannen (Ortega u. Aguillo 2010, 337). Da es sich bei der Untersuchung jedoch um eine Webseite und keinen Suchdienst handelt, müssen die Ergebnisse nicht für EconBiz gelten.

Generell kann man annehmen, dass Nutzer, die auf der Homepage starten EconBiz wahrscheinlich eher kennen werden und auch explizit aufgesucht haben, um eine Recherche zu starten, wohingegen der Einstieg auf Unterseiten und über eine Suchmaschine vermutlich meist unbeabsichtigt ist. Daher ist anzunehmen, dass die Nutzer, die auf der Homepage starten intensiver in EconBiz suchen, als die Nutzer, die dafür eine andere Suchmaschine

verwenden. Für letztere ist EconBiz vermutlich der Endpunkt der Suche, an dem die Ergebnisse „eingesammelt“ werden.

Um zu ermitteln, ob und wie das Suchverhalten vom Einstiegspunkt der Suche in EconBiz beeinflusst wird, wurden verschiedenen Parametern erhoben. Es zeigten sich eine ganze Reihe signifikanter Unterschiede gegenüber der Gesamtverteilung bei jedem Einstiegspunkt. Betrachtet wurden nur drei verschiedene Einstiegspunkte: der Start auf einem Einzeltreffer, auf einer Suchergebnisseite oder auf der Homepage. Weitere Seiten, wie etwa Unterseiten und Hilfeseiten, sind eher die Ausnahme und wurden daher auch nicht mitbetrachtet. Aus der Stichprobe starten 1.819 von 1.825 Sessions auf einer der drei betrachteten Seitentypen.

8.2 Einstiegspunkte

Der Aufruf der ersten Transaktion einer Session in der Stichprobe enthielt zu 53,92 % einen Einzeltreffer, zu 32,82 % eine Suchergebnisseite und zu 12,12 % die Startseite. Damit unterscheidet sich die Stichprobe deutlich von der Gesamtheit der Sessions, denn dort liegt die Verteilung bei 88,32 % auf Einzeltreffern, 8,99 % auf Suchergebnisseiten und nur 1,79 % auf der Startseite. Der Grund für diese Abweichung wird vermutlich in der Herausnahme der Sessions mit nur einem Seitenaufruf (Bounce-Sessions) liegen, die zum überwiegenden Anteil Einzeltreffer aufgerufen haben (s. 5.3).

8.3 Herkunft

Um festzustellen aus welchem Kontext die Nutzer in Econbiz landen, wurde in der Stichprobe eine Einstiegstaktiken vergeben. Es gab insgesamt 6 verschiedene Arten, auf die die Nutzer nach EconBiz gelangen konnten. Im Folgenden wird betrachtet, welche wie häufig vorkamen und woher die Nutzer meist kamen.

8.3.1 Ankunft über Suchmaschinen

80,82 % der Nutzer in der Stichprobe nutzten den Weg über eine Suchmaschine, um in EconBiz anzukommen. Genauer gesagt nutzten 78,46 % der Nutzer Google für den Einstieg in EconBiz, das sind 97,08 % der Nutzer, die eine Suchmaschine für den Einstieg nutzten. Nutzer, die von einer Suchmaschine aus in EconBiz landeten, kamen zu 66,58 % auf eine Einzeltrefferseite und zu 28,14 % auf eine Suchergebnisseite an. Weitere 4,68 % nutzten eine Suchmaschine, um auf die EconBiz Startseite zu gelangen. In letzteren Fällen kann wahrscheinlich davon ausgegangen werden, dass die Nutzer die feste Absicht hatten, EconBiz aufzusuchen.

8.3.2 Ankunft über zbw.eu

11,01 % der Nutzer in der Stichprobe kamen über die Webseite der ZBW nach EconBiz. 91,04 % der Nutzer, die über die ZBW-Webseite kamen, nutzten für den Einstieg in EconBiz den angebotenen Suchschlitz. D.h. das auch 91,04 % der Nutzer als erstes eine Suchergebnisseite aufgerufen hatten. Weitere 8,9 % riefen die Homepage auf.

8.3.3 Ankunft über Multiplikatoren

5,4 % der Nutzer in der Stichprobe kamen über Multiplikatoren nach EconBiz. Von diesen kamen alle auf der Homepage von EconBiz an. 69,69 % der Nutzer, die über Multiplikatoren kamen, benutzten DBIS (das Datenbank-Infosystem). Sie waren also auf der Suche nach einer passenden Datenbank für ihre Recherche.

Alle weiteren Wege nach EconBiz zu gelangen, kamen nur vereinzelt in der Stichprobe vor. Erwähnenswert ist noch, dass in 10 Fällen Verlinkungen auf den Seiten von wirtschaftsdienst.eu und intereconomics (den von der ZBW herausgegebenen Zeitschriften) Nutzer direkt auf eine Suchergebnisseite mit den Publikationen des Autors führte, dessen Artikel sie zuvor auf der Seite der Zeitschrift angesehen hatten.

8.4 Absicht oder „Hineinstolpern“?

Von Smiths Suchtaktiken wurden drei verschiedenen Typen genutzt um zu unterscheiden, welche Strategien die Nutzer für den Einstieg in EconBiz gewählt haben: direkt über eine Suchmaschine zum Ziel „teleportieren“ (*Teleport*), einen Anbieter (*Provider*) der gewünschten Inhalte aufsuchen oder eine bekannte Adresse direkt in den Browser eingeben (*URL*) (vgl. (Smith 2012, 10f.)). Sowohl für *Provider* als auch für *URL* ist die Bekanntheit des Anbieters eigentlich vorauszusetzen. Da viele Nutzer jedoch über andere Bibliotheken den Sucheinstieg fanden (z.B. DBIS), die so eigentlich nur Zugang zum Informationsanbietern gegeben haben, wurden in dieser Untersuchung auch diese zur Strategie *Provider* hinzugezählt.

Dahingegen gehören zu *Teleport* jene Fälle, wo die Nutzer mitten in die Ergebnisse hineinspringen (also von einer Suchmaschine aus eine Suchergebnisseite oder einen Einzeltreffer in EconBiz aufrufen). Nutzer, die den Suchschlitz auf der Webseite der ZBW nutzten, wurden auch zu *Provider* hinzugerechnet. Zwar ist nicht bekannt, wie sie auf die ZBW-Webseite gelangten, jedoch kann wahrscheinlich davon ausgegangen werden, dass sie zumindest die Absicht hegten, einen passenden Informationsanbieter für ihr Informationsbedürfnis aufzusuchen.

Rechnet man dann die Sucheinstiege so zusammen, kommt man auf 77 % wahrscheinlich unabsichtliche Sucheinstiege in EconBiz, gegenüber 23 % wahrscheinlich absichtlichen Sucheinstiegen. Bei einem Großteil der Nutzer kann also nicht unbedingt davon ausgegangen werden, dass sie absichtlich eine Suche in EconBiz machen wollten, oder dass sie EconBiz überhaupt schon kennen. Auch bei den Nutzern, die über Multiplikatoren in EconBiz eingestiegen sind, ist nicht vorauszusetzen, dass sie EconBiz schon kennen, jedoch eher, dass sie explizit eine spezielle Datenbank nutzen wollten. Die folgenden Abschnitte beleuchten, was die Nutzer als erstes taten, wenn sie wahrscheinlich unabsichtlich in EconBiz ankamen.

8.4.1 Erste Aktionen nach Hineinstolpern

In der zweiten Aktion einer Session, die durch die Taktik *Teleport* begann, klickten 26,6 % der Nutzer auf die übergeordnete Publikation eines Treffers auf einer Suchergebnisseite oder einer Einzeltrefferseite. 11,1 % wählten einen Treffer auf einer Suchergebnisseite in EconBiz aus, die sie über die Suchmaschine gefunden haben. Weitere 9,53 % kamen sofort wieder durch einen weiteren Klick in einer Suchmaschine auf eine andere Seite in EconBiz (Rebounding). Noch einmal 9,25 % klickten auf den Autornamen der über die Suchmaschine gefundenen Seite und noch 8,75 % benutzten den Suchschlitz in EconBiz für die Eingabe einer Suchanfrage. In 5,91 % der Fälle wurde in der zweiten Transaktion genau dieselbe Seite erneut aufgerufen, die bereits in der ersten Transaktion geladen wurde, entweder durch erneute Auswahl in der Suchmaschine oder durch technisch bedingte Reloads. Zählt man die Chaining-Taktiken zusammen, machen sie allein 49,08 % der Aktionen in der zweiten Transaktion aus. Natürlich unterscheidet sich das Verhalten der Nutzer, je nachdem, ob sie auf einer Suchergebnisseite oder einer Einzeltrefferseite in EconBiz angekommen waren.

8.4.1.1 Ankunft auf einem Einzeltreffer über eine Suchmaschine

Nutzer, die auf einem Einzeltreffer ankamen machten zwei Drittel der Nutzer aus, die über die Suchmaschinen kamen. Von den Nutzern, die auf einem Einzeltreffer landete, klickten 36,62 % zunächst auf die übergeordnete Publikation des Treffers. Weitere 10,27 % auf einen Autornamen und 10,07 % wechselten für die Suche in den Suchschlitz von EconBiz. Rebounding kam in weiteren 8,95 % vor, und Reloads in 8,55 %. Hier machen die Chaining-Taktiken zusammengenommen sogar fast zwei Drittel der ausgeführten Aktionen aus (65,71 %). 4,88 % der Nutzer klickten auf die rechts angebotenen „ähnlichen Treffer“. 1,4 % wählten dabei die Dublette des bereits gefundenen Treffers aus. Auch wurde häufiger auf „ähnliches Autorenspektrum“ geklickt, als auf „ähnliches Themenspektrum“.

8.4.1.2 Ankunft auf einer Suchergebnisseite über eine Suchmaschine

28,14 % der Nutzer kamen durch eine Suchmaschine auf einer Suchergebnisseite von EconBiz an. 37,68 % von diesen Nutzern wählten dann einen Treffer auf dieser gefundenen Suchergebnisseite aus. In 11,83 % wurde dieselbe Suchergebnisseite noch einmal geladen (Reloads oder Reselect). Davon könnten einige Fälle durch das Umsortieren der Ergebnisse entstanden sein, was jedoch nur in seltenen Fällen im Logfile nachweisbar ist (s. 6.1.3). 10,14 % wählten eine weitere Seite in einer Suchmaschine aus (Rebounding). 9,66 % blätterten auf der gefundenen Seite weiter und 7 % klickten auf einen Autornamen in der Suchergebnisliste und suchten so nach den Publikationen dieses Autors. 5,55 % schickten eine Anfrage über den Suchschlitz von EconBiz ab.

8.5 Verhalten nach Einstiegspunkten

Nachdem betrachtet wurde, wie viele Nutzer wahrscheinlich absichtlich nach EconBiz kamen und wo sie ankommen, wird jetzt betrachtet, wie sich die Sessions in verschiedenen Parametern unterscheiden, je nachdem auf welcher Seite die Nutzer in EconBiz einstiegen. Betrachtet wurden die Unterschiede bei den Sessionlängen und Dauern, dem Bouncing-Verhalten, dem Weiterblättern-Verhalten, der Nutzung des Suchschlitzes, den Reformulierungs- und Chaining-Taktiken, der Filternutzung sowie dem gemessenen Sucherfolg.

8.5.1 Sessionlänge

Bei Sessions, die auf einem Einzeltreffer starten, wird erwartet, dass die Länge deutlich kürzer ist, als bei anderen Sessions, da angenommen werden kann, dass die Suche in einer anderen Suchmaschine stattgefunden hat, und in EconBiz beendet wird. Die Anzahl der Seitenaufrufe zwischen Sessions, die auf einer Suchergebnisseite starten und einer Homepage sollte sich nicht deutlich unterscheiden, wahrscheinlich nur ein Seitenaufruf mehr, da die Homepage zunächst geladen wurde, statt in einer anderen Suchmaschine oder auf der ZBW-Webseite anzufragen.

Die Dauer der Session sollte mit den Seitenaufrufen zusammenhängen. Also ist zu erwarten, dass Sessions, die auf einem Einzeltreffer starten auch zeitlich deutlich kürzer ausfallen, als die anderen Sessions.

Die Länge der Session wurde in zwei Merkmalen ermittelt: die Anzahl der Seitenaufrufe und die zeitliche Dauer zwischen erster und letztem Seitenaufruf. Daneben wurde auch eine durchschnittliche Seitenaufruffrequenz errechnet. Tabelle 8.1 zeigt die Werte nebst p-Werte aus den Signifikanztests. Das alpha-Niveau beträgt in allen Fällen 0,05.

Merkmal	Einzeltr.	Sucherg.	Home	Signifikanz
Anzahl der Sessions	986	611	222	-
ØSeitenaufrufe	5,65	5,68	9,86	Ja: f-Wert=34,22 > 3
ØSeitenaufrufe ohne Return/Reload	4,09	4,3	8,22	-
ØSessiondauer	361 s	450 s	668 s	Ja: f-Wert=12,56 > 3
ØFrequenz der Seitenaufrufe	59 s	51 s	61 s	Nein: f-Wert:1,6 < 3

Tab. 8.1: Längen und Dauern der Sessions nach Einstiegspunkt

Es fällt auf, dass sich eigentlich nur die Sessions, die auf der Homepage starten, von den anderen unterscheiden. Sie sind sowohl in zeitlicher Hinsicht länger, als auch mit häufigeren Seitenaufrufen verbunden. Die Seitenaufrufe ohne Reload und Zurückgeh-Aktionen wurden noch hinzugezogen, um auszuschließen, dass Verfälschungen durch technisch bedingte Seitenaufrufe aufgetreten sind. Die Abstände haben sich dadurch jedoch kaum verändert. Die Unterschiede in der Seitenaufruffrequenz sind nicht signifikant. Hier konnte also kein Unterschied festgestellt werden.

8.5.2 Bouncing

Die Tabelle 8.2 zeigt die Werte der Hineinsprünge aus einer Suchmaschine in EconBiz, aufgeteilt nach Einstiegspunkten. Die Durchschnitte wurden berechnet sowohl über alle Sessions mit dem Einstiegspunkt X, als auch nur über die Sessions, in denen das Merkmal überhaupt auftrat, um eine Intensität des Verhaltens zu ermitteln.

Merkmal	Einzeltr.	Sucherg.	Home	Signifikanz
Anzahl der Sessions	986	611	222	-
% Sessions mit Bouncing	99,8 %	68,25 %	4,5 %	Ja: $p= 9,19 \times 10^{-168}$
ØBounces	1,26	0,24	0,05	Ja: f-Wert=178,67 > 3
Øwenn Bounces ≥ 1	1,26	1,25	1,2	Nein: f-Wert=0,04 < 3

Tab. 8.2: Bouncing-Verhalten nach Einstiegspunkten

Am seltensten wird in Sessions, die auf der Homepage starten aus anderen Suchmaschinen in EconBiz hineingesprungen. Das wird daran liegen, dass die Nutzer, die auf der Homepage starten, EconBiz vermutlich absichtlich und explizit für ihre Recherche aufgesucht haben, anstatt eine andere Suchmaschine zu verwenden.

Der Unterschied zwischen dem Start auf Einzeltreffer und Suchergebnisseiten wird sich daraus ergeben, dass ein guter Anteil der Nutzer, die auf einer Suchergebnisseite einstiegen, den Suchschlitz auf der Webseite der ZBW genutzt hatte, und daher auch zu den Nutzer gerechnet werden kann, die EconBiz vermutlich explizit für ihre Suche aufgesucht haben.

8.5.3 Weiterblättern

Das Weiterblättern auf den Suchergebnisseiten sollte eigentlich unabhängig vom Einstiegspunkt der Session erfolgen. Sowohl durch Suchanfragen im Suchschlitz, als auch durch Chaining werden Suchergebnisseiten erzeugt, die man intensiv oder weniger intensiv durchblättern kann. Die Tabelle 8.3 zeigt die Werte der Aktionen, in denen die Nutzer auf einer Suchergebnisseite weitergeblättert haben, aufgeteilt nach Einstiegspunkten.

Merkmal	Einzeltr.	Sucherg.	Home	Signifikanz
Anzahl der Sessions	986	611	222	-
% Sessions mit Weiterblättern	37,63 %	20,79 %	31,08 %	Ja: $p = 3,51 \times 10^{-8} < 0,05$
Øweitergeblättert	0,38	0,72	1,42	Ja: f-Wert=19,74 > 3
Øwenn weitergeblättert ≥ 1	2,7	3,49	3,35	Nein: f-Wert=2,95 < 3

Tab. 8.3: Weiterblätter-Verhalten nach Einstiegspunkten

Die Unterschiede sind beim Weiterblättern auf Suchergebnisseiten nicht sehr groß, aber dennoch vorhanden. Das Verhalten beim Weiterblättern ist signifikant abhängig vom Einstiegspunkt der EconBiz-Suche. Interessanterweise ist der Anteil der Sessions, in denen weitergeblättert wurde bei den Einzeltreffern am größten. Dies mag durch die Wahl der Stichprobe gekommen sein, da die meisten Sessions, die auf einem Einzeltreffer gestartet sind, ja meist keine weiteren Seiten aufgerufen haben. Bei der Intensität des Weiterblätterns konnte kein signifikanter Unterschied nachgewiesen werden.

8.5.4 Nutzung des Suchschlitzes

Es ist zu erwarten, dass Sessions, die auf einem Einzeltreffer starten, deutlich seltener den Suchschlitz von EconBiz nutzen, als z.B. die Nutzer, die auf der Homepage starten, denn der Einzeltreffer kann bereits das (evtl. befriedigende) Ergebnis einer Suche in einer anderen Suchmaschine sein. Die Anzahl der Suchanfragen, die über den Suchschlitz abgeschickt werden, könnte sich auch dahingehend verändern, dass Nutzer, die auf der Homepage starten EconBiz vermutlich absichtlich aufgesucht haben für eine Recherche und den Suchschlitz daher innerhalb der Sessions intensiver nutzen, als andere.

Die Tabelle 8.4 zeigt die Werte der Aktionen, in denen die Nutzer auf eine Suchanfrage über den Suchschlitz abgeschickt hat (sowohl die ersten, als auch die reformulierten), aufgeteilt nach Einstiegspunkten.

Der Suchschlitz wird in drei Viertel der Sessions, die auf einem Einzeltreffer starten, gar nicht verwendet. Dahingegen wird er in jedem Fall verwendet, wenn die Nutzer auf der Homepage starten. Auch bei den Sessions, die auf Suchergebnisseiten landen, wird in über der Hälfte

Merkmal	Einzeltr.	Sucherg.	Home	Signifikanz
Anzahl der Sessions	986	611	222	-
% Sessions mit Suchanfrage	24,04 %	39,12 %	100 %	Ja: $p=1,37 \times 10^{-37} < 0,05$
ØAnzahl Suchanfragen	0,38	0,72	1,42	Ja: f-Wert=105 > 3
ØAnzahl Suchanfragen wenn ≥ 1	2,7	3,49	3,35	Nein: f-Wert=0,68 < 3

Tab. 8.4: Nutzung des Suchschlitzes nach Einstiegspunkten

der Fälle der Suchschlitz nicht genutzt, wenn dann jedoch genauso intensiv, wie in den Sessions, die auf der Homepage beginnen. Da Sessions, die auf einer Suchergebnisseite beginnen sowohl absichtlich, als auch unabsichtlich EconBiz aufgesucht haben könnten, liegt es nahe zu vermuten, dass die Sessions, in denen der Suchschlitz verwendet wurde, jene sind, die von der ZBW-Webseite einstieg (dann wäre bereits die erste Aktion eine Suchanfrage). Dies liegt wiederum nahe, dass Nutzer, die eine Suchergebnisseite in einer anderen Suchmaschine aufgerufen haben, kaum dazu neigen, dann den Suchschlitz in EconBiz zu verwenden.

8.5.5 Suchtypen

Die Verwendung von thematischer oder Known-Item-Suche sollte sich eigentlich nicht nach Einstiegspunkten unterscheiden. Eventuell ist es schwieriger eine Known-Item-Suche nachzuweisen, wenn diese eigentlich bereits vor dem Einstieg in EconBiz stattgefunden hatte. Daher wurde erwartet, dass bei Suchen, die auf Einzeltreffern starteten, besonders selten eine Known-Item-Suche in der Session erfolgte.

Die Tabelle 8.5 zeigt die Werte der Aktionen, in denen der Nutzer auf eine Known-Item-Suchanfrage über den Suchschlitz abgeschickt hat (nur die initialen Anfragen), aufgeteilt nach Einstiegspunkten.

Merkmal	Einzeltr.	Sucherg.	Home	Signifikanz
Anzahl der Sessions	986	611	222	-
% mit Known-Item-Suche	14,5 %	20,62 %	42,34 %	Ja: $p=0,00015 < 0,05$
ØAnzahl Suchanfragen	0,19	0,28	0,52	Ja: f-Wert=18,79 > 3
ØAnzahl Suchanfragen wenn ≥ 1	1,32	1,33	1,23	Nein: f-Wert=0,17 < 3

Tab. 8.5: Suchtypen nach Einstiegspunkten

Es zeigt sich auch hier, dass Nutzer, die auf der Homepage einstieg am häufigsten eine Known-Item-Suche in der Session ausgeführt haben. Jedoch unterscheiden die Einstiegspunkte sich nicht nach der Intensität, mit der verschiedenen Known-Items in den Sessions gesucht wurden. Wie erwartet ließen sich bei Einzeltreffer-Startern am seltensten Known-Item-Suchen erkennen, was wahrscheinlich mit der geringen Suchschlitznutzung überhaupt zusammenhängt.

8.5.6 Reformulierungstaktiken

Der Einsatz von Reformulierungstaktiken hängt mit der Intensität der Suche zusammen. Diese sollten also besonders häufig in Sessions vertreten sein, die auf der Homepage begannen, da diese Nutzer EconBiz explizit für eine Recherche aufgesucht haben werden.

Die Tabelle 8.6 zeigt die Werte der angewendeten Reformulierungstaktiken, wenn Nutzer eine Suchanfrage über den Suchschlitz veränderten, aufgeteilt nach Einstiegspunkten.

Merkmal	Einzeltr.	Sucherg.	Home	Signifikanz
Anzahl der Sessions	986	611	222	-
% Sessions mit Reform.	9,23 %	17,68 %	44,15 %	Ja: $p=1,67 \times 10^{-6} < 0,05$
ØAnzahl Reform.	0,16	0,38	1,14	Ja: f-Wert=74,48 > 3
ØAnzahl Reform. wenn ≥ 1	1,77	1,82	2,32	Ja: f-Wert=3,43 > 3

Tab. 8.6: Reformulierungsverhalten nach Einstiegspunkten

Über alle Sessions gesehen, werden Reformulierungstaktiken in 16 % der Sessions der Stichprobe verwendet. Es ist jedoch eine Abhängigkeit vom Einstiegspunkt gegeben. Besonders wenige Reformulierungstaktiken enthalten Sessions, die auf einem Einzeltreffer starten. Das erstaunt nicht, wenn man ja sieht, dass knapp ein Viertel überhaupt den Suchschlitz verwendet. Nutzer, die auf einer Suchergebnisseite starten, reformulieren eine Anfrage fast durchschnittlich oft, wohingegen die Nutzer, die auf der Homepage starten deutlich häufiger ihre Suchanfragen noch einmal verändern. Auch dies weist also daraufhin, dass die Nutzer, die auf der Homepage starten intensiver in EconBiz suchen, als die anderen.

8.5.7 Chaining-Taktiken

Chaining-Taktiken bieten sich in erster Linie an, wenn man auf einem Einzeltreffer oder einer Suchergebnisseite einsteigt, von der Homepage aus sind sie gar nicht möglich. Spätestens jedoch ab der ersten Suchergebnisseite sind sie es, deshalb sollte der Unterschied hier nicht ganz so deutlich ausfallen. Die Tabelle 8.7 zeigt die Werte der angewendeten Chaining-Taktiken, aufgeteilt nach Einstiegspunkten.

Merkmal	Einzeltr.	Sucherg.	Home	Signifikanz
Anzahl der Sessions	986	611	222	-
% Sessions mit Chaining	82,86 %	25,7 %	14,41 %	Ja: $p=3,7 \times 10^{-81} < 0,05$
ØAnzahl Chaining	1,11	0,32	0,23	Ja: f-Wert=226 > 3
ØAnzahl Chaining wenn ≥ 1	1,26	1,19	1,56	Nein: f-Wert=2,38 < 3

Tab. 8.7: Chaining-Verhalten nach Einstiegspunkt

Auch die Anwendung von Chaining-Taktiken ist signifikant abhängig vom Einstiegspunkt der Suche. Während über alle Sessions gesehen 54 % mindestens eine Chaining-Taktik enthalten, ist der Anteil bei den Sessions, die auf einem Einzeltreffer starten deutlich höher, dafür bei den anderen beiden Einstiegspunkten deutlich niedriger. Statt also den Suchschlitz zu verwenden, klicken sich die Nutzer, die auf einem Einzeltreffer ankommen meist eher durch, wohingegen die Nutzung des Suchschlitzes besonders bei den Nutzern, die auf der Homepage starten, intensiver ist. Die Intensität des Chainings unterscheidet sich bei den verschiedenen Einstiegspunkten dabei nicht, wie bei der Suchschlitznutzung. Chaining wurde generell eher selten innerhalb derselbe Session wiederholt genutzt.

8.5.8 Filternutzung

Die Filternutzung wird mit der Intensität, mit der sich mit EconBiz auseinandergesetzt wird, zusammenhängen. Nutzer, die auf einem Einzeltreffer starten, werden durch die meist vorher erfolgte Suche in einer anderen Suchmaschine weniger häufig die Filter nutzen, als die anderen. Die Tabelle 8.8 zeigt die Werte der Filtervorgänge, aufgeteilt nach Einstiegspunkten.

Merkmal	Einzeltr.	Sucherg.	Home	Signifikanz
Anzahl der Sessions	986	611	222	-
% Sessions mit Filtern	6,19 %	14,89 %	26,13 %	Ja: $p=0,01 < 0,05$
ØAnzahl Filter	0,11	0,31	0,66	Ja: $f\text{-Wert}=28,18 > 3$
ØAnzahl Filter wenn ≥ 1	1,63	2,11	1,76	Nein: $f\text{-Wert}=0,05 < 3$

Tab. 8.8: Filterverhalten nach Einstiegspunkten

Auch bei der Filternutzung bestätigt sich: Nutzer, die auf einem Einzeltreffer ankommen, tun wenig mehr, als sich per Chaining durchzuklicken. Nur in 6 % der Sessions, die auf einem Einzeltreffer starten, wird mindestens einmal ein Filter gesetzt. Wohingegen in fast einem Viertel der Sessions, die auf der Homepage starteten mindestens einmal die Filter genutzt wurden. Die Intensität der Filternutzung unterscheidet sich nicht.

8.5.9 Sucherfolg

Erfolge zeigen die Klicks auf Einzeltreffer auf Suchergebnisseiten, oder das erfolgreiche Ranking eines Known-Items auf Rang 1 an. Die Tabelle 8.9 zeigt die Werte der registrierten Erfolg in einer Session, aufgeteilt nach Einstiegspunkten.

Die registrierten Erfolge waren ebenfalls abhängig vom Einstiegspunkt der Suche. Da in Sessions, die auf einem Einzeltreffer starteten, weniger Suchergebnisseiten aufgerufen werden, als in den anderen Sessions, hatten sie von vornherein weniger Chance, einen Erfolg nach der Definition in dieser Arbeit zu erhalten. Besonders viele Sessions mit mindestens einem Erfolg

Merkmal	Einzeltr.	Sucherg.	Home	Signifikanz
Anzahl der Sessions	986	611	222	-
% Sessions mit Erfolg	27,99 %	56,79 %	64,41 %	Ja: $p = 3,5 \times 10^{-17} < 0,05$
ØAnzahl Erfolge	0,42	0,89	1,46	Ja: f-Wert=69,44 > 3
ØAnzahl Erfolge wenn ≥ 1	1,5	1,56	2,12	Ja: f-Wert=6,07 > 3

Tab. 8.9: Sucherfolg nach Einstiegspunkten

finden sich wiederum bei den Sessions, die auf der Homepage starten. Hier ist gleichzeitig auch der Anteil der Sessions hoch, die mindestens eine Known-Item-Suche enthalten, weshalb einige Erfolge bestimmt nicht durch Klicks sondern durch erfolgreiches Ranking entstanden sind. Auch die Unterschiede zwischen den Einstiegspunkten bei der Intensität von Erfolg ist signifikant. In Sessions, die auf der Homepage starten, wurde häufiger ein Klick erkannt (oder ein Known-Item auf Rang 1 gerankt), als in den anderen Sessions.

8.6 Fazit zu den Einstiegspunkten

Die Nutzer, die auf der Homepage starten, scheinen sich tatsächlich intensiver mit den Suchergebnissen auseinanderzusetzen und die Suche häufiger zu verwenden, als Nutzer die auf einer Unterseite gestartet sind. Sie verwenden den Suchschlitz insgesamt häufiger, reformulieren ihre Suchanfragen öfter, und nutzen die Facettenfilter öfter, als die anderen Nutzer. Dafür verwenden sie deutlich seltener die Chaining-Taktiken. 85 % der Sessions, die auf der Homepage starten verwenden nicht einmal eine Chaining-Taktik. Der Unterschied zu den Sessions, die auf einer Suchergebnisseite starten, ist stets nicht ganz so deutlich, wie die Unterschiede zu den Sessions, die auf einem Einzeltreffer starten, da ein großer Anteil dieser Nutzer über die ZBW-Homepage absichtlich EconBiz aufgesucht haben wird.

Die Nutzer, die auf der Homepage starten, verwenden auch mehr unterschiedliche Taktiken, als andere und verbringen insgesamt mehr Zeit in EconBiz, als die anderen. In der Stichprobe sind die Sessions fast doppelt so lang, als bei Nutzern, die auf einem Einzeltreffer einstiegen. Sie riefen im Schnitt auch doppelt so viele Seiten auf.

Weiterhin ist auffällig, dass in fast keiner Session, die auf der Homepage startet, irgendwann auch von einer anderen Suchmaschine in EconBiz hineingesprungen wird (Bouncing). Dies geschah nur in 10 Sessions. Aus der interpretativen Analyse heraus zeigt sich hier dann, dass die Nutzer mit der Suche in EconBiz nicht erfolgreich waren, und die gesuchte Publikation dann über Google gesucht und gefunden hatten.

Das Suchverhalten in Sessions, die auf einem Einzeltreffer starten, legt einen deutlichen Schwerpunkt auf Chaining-Taktiken. Das Durchklicken ist hier der wesentliche Weg, wie sich die Nutzer mit EconBiz auseinander setzen. Die eigentliche Suche mit Suchanfragen findet

jedoch in einer anderen Suchmaschine statt und EconBiz wird eher selten als Suchmaschine verwendet, sondern mehr als Website, die Content anbietet. Somit bestätigte sich die Vermutung, dass die Ergebnisse von Ortega u. Aguillo (2010) sich nicht auf EconBiz übertragen lassen (s. 8.1).

Kapitel 9

Fazit

9.1 Zusammenfassung

Durch die Logfile-Analyse aus einem Jahr EconBiz Nutzung konnte ein besseres Bild über das Nutzungsverhalten der EconBiz-Nutzer erlangt werden. Es zeigte sich:

1. dass die Nutzer meist über eine andere Suchmaschine auf Einzeltreffer-Seiten in EconBiz ankommen.
2. dass die Nutzer häufig auch gleich wieder abspringen und nicht mehr wiederkommen.
3. dass Nutzer, die über die Homepage oder die Seiten der ZBW kommen, EconBiz deutlich intensiver für eine Recherche nutzen.
4. dass die Nutzer in vielen Sessions nie den Suchschlitz verwenden, sondern sich durch EconBiz durchklicken.
5. dass die Nutzer besonders häufig auf die übergeordnete Publikation eines Einzeltreffers klicken, was wohl eine transaktionsorientierte Handlung sein dürfte.
6. dass die Nutzer eher selten ihre Suchanfragen reformulieren.
7. dass die Nutzer ihre Suchanfragen oft gleich mehrmals reformulieren, wenn sie es denn tun und dabei häufig unterschiedliche Strategien verwenden.
8. dass die Nutzer bei thematischen Suchen eher dazu neigen, die Suchbegriffe zu wechseln, als die Suchanfragen zu verlängern.
9. dass EconBiz wahrscheinlich häufig das Ziel von Known-Item-Suchen ist.
10. dass die Nutzer bei Known-Item-Suchen am häufigsten den vollständigen Titel verwenden und diesen im Falle einer Reformulierung verkürzen.
11. dass die Nutzer bei Known-Item-Suchen bei nötigen Reformulierungen die Suchanfragen lieber verlängern mit weiteren Angaben, als die Angaben zu wechseln.

12. dass die Nutzer häufig genau die Publikation suchen, die sie eigentlich bereits über eine andere Suchmaschine in EconBiz gefunden hatten.
13. dass die Nutzer die Filter eher selten einsetzen, und dann häufig nach Themen/Schlagnworte filtern
14. dass die Nutzer häufiger bei einer thematischen Suche filtern, als bei einer Known-Item-Suche.
15. dass die Nutzer gerne gleich zweimal nach Themen/Schlagnworte filtern: auf deutsch und englisch.

Die Erkenntnisse aus dieser Untersuchung decken sich im Wesentlichen auch mit den Ergebnissen aus anderen bisherigen Untersuchungen in ähnlichen Bereichen. Aspekte, die dort nicht auftauchten, waren bei Known-Item-Suchen das Wiedersuchen von bereits gefundenen Treffern, was sich damit erklärt, dass das Zusammenwirken von Websuche und Recherche in Katalogen oder Datenbank bei den Untersuchungen von Known-Item-Suchen nicht betrachtet wurde.

Es konnten durch die Logfile-Analyse auch einige Punkte erkannt werden, an denen die Nutzer auf Probleme stoßen:

Labeling-Missverständnis Bei der Benennung des Links in der Publikationsform (Subkategorie) auf den Einzeltrefferseiten kommt es vermutlich oft zu einem Missverständnis. Die Nutzer vermuten wahrscheinlich Zugang oder Informationen zum Zugang oder zur übergeordneten Publikation hinter „Artikel in Zeitschrift“. Dies zeigte sich daran, dass eigentlich nie ein Treffer ausgewählt wurde, wenn diese Funktion genutzt wurde, und dass häufig danach noch einmal auf den Link auf dem Zeitschriftentitel geklickt wurde. Dieses Label taucht natürlich nicht bei jedem Einzeltreffer auf, das Metadatum muss vorhanden sein. Dennoch stellt sich die Frage, wie sinnvoll die Funktion eigentlich ist. Aus welchem Grund würde ein Nutzer alle Zeitschriftenartikel auf einen Blick sehen wollen?

Umstellen der Feldeinschränkung Die automatische Verstellung der Feldeinschränkung nach Chaining-Klicks macht zwar transparent, was für Ergebnisse angezeigt werden. Allerdings nehmen die Nutzer die Verstellung anscheinend oft nicht wahr (vermutlich weil sie sie nicht aktiv vornehmen). Kommt es dann zu einem Wechseln in den Suchschlitz so wurden häufig 0 Treffer oder sehr wenige Ergebnisse erzielt. Die Feldeinschränkung automatisch zu verstellen sollte besser unterbleiben und die Transparenz dessen, was beim Chaining-Klick passiert auf andere Weise dargestellt werden. Zwar

kam es nicht sehr häufig vor, dass nach Chaining noch einmal in den Suchschlitz gewechselt wird, jedoch ist das Potential, dass dieser Fehler auftreten kann groß, in Anbetracht der Tatsache, dass Chaining die häufigste Taktik ist. Dagegen ist die aktive Nutzung dieser Funktion nur vereinzelt aufgetreten. In Transaktionen, in denen die Nutzer aktiv die Feldeinschränkung veränderten, wurde meist nur eine durch Chaining verstellte Beschränkung zurückgesetzt. Vielleicht ist diese Funktion tatsächlich für die aktive Nutzung in der Standardoberfläche so unbedeutend, dass man sie allein für die Erweiterte Suche anbieten sollte.

9.2 Grenzen der Untersuchung

Neben den bereits unter 4.1 aufgeführten generellen Beschränkungen der Aussagefähigkeit von Logfile-Analysen gibt es noch ein paar weitere mögliche Fehlerquellen bei dieser Untersuchung. Zum einen ist bei einer intellektuellen Klassifizierung von Suchanfragen immer ein Interpretationsspielraum gegeben, der bei unterschiedlichen Untersuchenden zu abweichenden Ergebnissen führen kann. Auch bei der Ermittlung, ob ein Known-Item indexiert war oder nicht, können einige vielleicht durch ungünstige Suchanfragen nicht gefunden worden sein. Generell ist die Stichprobe von nicht einmal 2.000 Sessions verhältnismäßig klein und daher könnten z.B. einige von Nutzern verwendete Taktiken nicht beobachtet worden sein. Durch die Zufallsauswahl sollte die Stichprobe jedoch zumindest für den Teil der Sessions repräsentativ sein, der mehr als eine Seite und mindestens eine Suchergebnisseite aufgerufen hat.

Konzeptionelle Kritik kann man im Nachhinein daran üben, dass man einerseits die in Known-Item-Suchanfragen verwendeten Elemente ermitteln wollte, und diese andererseits im Vorhinein fest definiert. Im Grunde ist dies ein Zirkelschluss. Ohne jedoch den Nutzer über seine Intention befragen zu können, ist eine solche operationale Definition von Known-Item-Suchen notwendig. Deutlich wurde in einem Fall, dass Nutzer unter Umständen auch Inhaltliches für eine Known-Item-Suche verwenden. Der Nutzer suchte nach einer bestimmten Zeitschriftenausgabe, die sich mit der Verwendung von Satellitenfotos in der Forschung befasste. Nur durch den Sessionverlauf und der Tatsache, dass er auch bibliographische Angaben für die Suche verwendete, konnte erkannt werden, dass auch die Suchanfrage „satellitenfotos“ eine Known-Item-Suche war, obwohl sie kein bibliographisches Datum enthielt. Es gibt also vermutlich eine Dunkelziffer an Known-Item-Suchen, die einen thematischen Zugang gewählt haben.

Weiterhin war das Auftreten von Information-Encountering-Effekten und die Dubletten-Problematik bei Known-Item-Suchen in der Datenstruktur der Datenbank nicht vorgesehen

und führte bei der Bewertung des Sucherfolgs zu Problemen. Letztlich wurde der sichtbare Klick über das Ranking gestellt, was die Daten zur Retrievaleffektivität in einigen Fällen negativ beeinflusst haben könnten. Die beiden Erfolgsbewertungen hätten getrennt beobachtet werden müssen.

Auch stellte sich im Verlauf des Projekts heraus, dass das gewählte Tool der relationalen Datenbank nicht für alle Fragestellungen das beste war. Zukünftige Untersuchungen sollten besser auf einen Aspekt des Suchverhaltens fokussieren, um auch die Datenstruktur entsprechend passend zu erhalten. In dieser ist es nicht ganz gelungen eine Struktur für alle Forschungsfragen passend zu gestalten.

9.3 Weitere Untersuchungen

Eine genauere Untersuchung mit anderen Mitteln als einer Logfile-Analyse, sollte überprüfen, wie verständlich die Zugangsinformationen und die Verfügbarkeitsprüfungsfunktion in EconBiz sind und wie sie genutzt werden. Gerade weil der Verdacht nahe liegt, dass ein Großteil der Sessions, in denen nur Chaining-Taktiken verwendet werden, allein versuchen zu eruieren, wie man auf die gefundene Publikation zugreifen kann. Es wäre interessant zu ermitteln, mit welchen Intentionen die Nutzer tatsächlich die Chaining-Klicks ausführen und was für Funktionen sie hinter den Links auf den Einzeltrefferseiten vermuten.

Da es leider nicht möglich war, zu erkennen, ob die Nutzer letztlich den Zugang gefunden haben, konnte kaum etwas über Erfolg oder Misserfolg ausgesagt werden. Eine Möglichkeit in Zukunft vielleicht ein besseres Bild über echte „Conversion“ bei Einzeltrefferaufrufen zu erhalten, wäre zu beobachten, ob überhaupt ein Online-Zugang vorhanden ist, und ob dieser auch frei ist. Die Nutzer der untersuchten Sessions im Logfile waren alle nicht direkt in der ZBW (die IP-Adressen wurden gefiltert). Somit ist eine direkte Conversion bei Print-Artikeln ausgeschlossen, ebenso wahrscheinlich bei zugangsbeschränkten Online-Medien. Die Einteilung der Einzeltrefferaufrufe in diese Stufen könnte ein verlässlicheres Bild, wie oft die Nutzer direkt über EconBiz an einen Artikel gekommen sein könn(t)en.

Die Frage, ob Nutzer, die Probleme haben ihr Informationsbedürfnis in EconBiz zu befriedigen, ein anderes Verhalten haben, als Nutzer, die erfolgreich sind, sollte mit anderen Methoden ermittelt werden. Der kritische Punkt ist die Ermittlung der Befriedigung des Informationsbedürfnisses. Erst in einem zweiten Schritt könnte dann versucht werden, die Beobachtungen auf den Logfile zu übertragen, ob dort ähnliche Muster erkennbar sind.

Da die Nutzer, wie sich gezeigt hat, am häufigsten nach Themen/Schlagworten filtern, könnte eine weitere Untersuchung ermitteln, welche Schlagworte in welchen Fällen gefiltert wurden.

Interessant wäre auch zu ermitteln, was die Nutzung der Facettenfilter auslöst, also welche Situation den Nutzer dazu anhält, die Facettenfilter zu nutzen.

Besonders interessant wäre eine Untersuchung, welche Rolle das Zusammenspiel der Nutzung von allgemeiner Websuche und EconBiz bei der Recherche für die Nutzer spielt und wie es abläuft.

Ein weiterer Bereich, der in dieser Untersuchung ausgeklammert wurde, wäre die Betrachtung der Sprachpräferenzen der Nutzer von EconBiz bei ihren Recherchen.

Literaturverzeichnis

Abhishek u. a. 2012

ABHISHEK, Vibhanshu ; FADER, Peter.S. ; HOSANAGAR, Kartik: Media exposure through the funnel: a model of multi-stage attribution. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2158421. Version:2012, Abruf: 2015-03-16

Aula u. a. 2010

AULA, Anna ; KHAN, Rehan M. ; GUAN, Zuwei: How does search behaviour change as search becomes more difficult? In: CHI 2010: Exploratory Search, 2010, S. S. 35–44

Bates 1990

BATES, Marcia J.: Where should the person stop and the information search interface start? In: Information Processing & Management 26 (1990), Nr. 5, S. 575–591

Boldi u. a. 2011

BOLDI, Paolo ; BONCHI, Francesco ; CASTILLO, Carlos ; VIGNA, Sebastiano: Query Reformulation Mining: Models, Patterns, and Applications. In: Information Retrieval 14 (2011), Nr. 3, S. 257–289

Borst 2012

BORST, Timo: Usage and impact of controlled vocabularies in a subject repository for indexing and retrieval. In: LIBER Quarterly 21 (2012), Nr. 3/4

Broder 2002

BRODER, Andrei: A taxonomy of web search. In: SIGIR Forum Bd. 36, 2002, S. 3–10

Bucklin u. Sismeiro 2009

BUCKLIN, Randolph E. ; SISMEIRO, Catarina: Click here for internet insight: advances in clickstream data analysis in marketing. In: Journal of interactive marketing 23 (2009), S. 35–48

Chapman u. a. 2013

CHAPMAN, Suzanne ; DESAI, Shevon ; HAGEDORN, Kat ; VARNUM, Ken ; MISHRA, Sonali ; PIACENTINE, Julie: Manually Classifying user search queries on an academic library web site. In: Journal of Web Librarianship 7 (2013), Nr. 4, S. 401–421

Ellis 1989

ELLIS, David: A behavioural approach to information retrieval system design. In: Journal of Documentation 45 (1989), Nr. 3, S. 171–212

Friedrichsen 2013

FRIEDRICHSEN, Merle: Die Sprache der Kunden sprechen: Eine Usability-Studie zu verständlichen Labels von der Literaturrecherche in EconBiz (Bachelorarbeit), HAW-Hamburg, Diplomarbeit, 2013

Genicus 2008

GENICUS, Research & C.: Evaluierung der Produkte und Serviceleistungen der Deutschen Zentralbibliothek für Wirtschaftswissenschaften Ergebnisse einer qualitativen Befragung von Nutzerinnen und Nutzern und von Fachreferentinnen und Fachreferenten. 2008

Huang u. Efthimiadis 2009

HUANG, Jeff ; EFTHIMIADIS, Efthimis N.: Analyzing and evaluating query reformulation strategies in web search logs. In: CIKM 09: Proceedings of the 18th ACM conference on Information and knowledge management, 2009, S. 77–86

Jansen u. a. 2007a

JANSEN, Bernard J. ; SPINK, Amanda ; BLAKELY, Chris ; KOSHMAN, Sherry: Defining a session on Web Search Engines. In: Journal of the American Society for Information Science and Technology 58 (2007), Nr. 6, S. 862–871

Jansen u. a. 2007b

JANSEN, Bernard J. ; ZHANG, Mimi ; SPINK, Amanda: Patterns and transitions of query reformulation during web searching. In: International Journal of Web Information Systems 3 (2007), Nr. 4, S. 328–240

Krebs 2013

KREBS, Matthias: Die known item search in Bibliothekskatalogen - Nutzerverhalten, Probleme, Lösungsmöglichkeiten. In: Bibliothek, Forschung und Praxis 38 (2013), Nr. 1, S. 91–102

Lee u. a. 2006

LEE, Jin H. ; RENEAR, Allen ; SMITH, Linca C.: Known-Item Search: Variations on a Concept. In: Proceedings of the American Society for Information Science and Technology 43 (2006), Nr. 1, S. 1–17

Lewandowski 2011

LEWANDOWSKI, Dirk: The retrieval effectiveness of search engines on navigational queries. In: ASLIB Proceedings 61 (2011), S. 354–363

Linek u. a. 2012

LINEK, Stephanie ; SCHAFRICK, Annela ; STEINHAGEN, Elisabeth ; SAGER, Jana: Labeling-Test zur Optimierung der Terminologie von EconBiz. November 2012

MacFarlane 2007

MACFARLANE, A.: Evaluation of web search for the information practitioner. In: ASLIB Proceedings 59 (2007), Nr. 4/5, S. 352–366

Neubert 2009

NEUBERT, Joachim: Statistik Volltextrecherchen in Econis. 2009

Nielsen 2005

NIELSEN, Jakob: Mental models for search are getting firmer. <http://www.nngroup.com/articles/mental-models-for-search/>. Version: May 2005, Abruf: 2015-03-10

Niu 2012

NIU, Xi: Beyond Text Queries and Ranked Lists: Faceted Search in Library Catalogs, University of North Carolina, Diss., 2012

Niu 2013

NIU, Xi: Faceted Search in Librarz Catalogs. In: PARK, Jung-Ran (Hrsg.) ; HOWARTH, Lynne C. (Hrsg.): New Directions in Information Organization. 2013, S. 173–208

Ortega u. Aguillo 2010

ORTEGA, José Luis ; AGUILLO, Isidro: Differences between web sessions according to the origin of their visits. In: Journal of Infometrics 4 (2010), S. 331–337

Ozmutlu u. a. 2009

OZMUTLU, Seda ; OZMUTLU, Huseyin C. ; SPINK, Amanda: Topic analysis and identification of queries. In: BERNARD J. JANSEN, Isak T. Amanda Spink S. Amanda Spink (Hrsg.): Handbook of research on web log analysis. Information Science references, 2009, S. 345–358

Petras u. a. 2014

PETRAS, Vivien ; STILLER, Juliane ; GÄDE, Maria: How we are searching cultural heritage? A qualitative analysis of search patterns and success in The European Library. In: Libraries in the digital age (LIDA) Proceedings 13 (2014)

Pianos 2014

PIANOS, Tamara: Online-Umfrage EconBiz März 2014. 2014

Rulik 2014

RULIK, Imke: Known-Item Suchanfragen im Discoverysystem beluga: Retrievaleffektivität und Empfehlungen, HAW-Hamburg, Diplomarbeit, 2014

Smith 2012

SMITH, Alastair G.: Internet search tactics. In: Online Information Reviews 36 (2012), Nr. 1, S. 7–20

Spink u. a. 2000

SPINK, Amanda ; JANSEN, Bernard J. ; OZMULTU, Cenk: Use of query reformulation and relevance feedback by Excite users. In: Internet Research 10 (2000), Nr. 4, S. 317–328

Stegemann u. Swanson 2010

STEGEMANN, Caroline ; SWANSON, Martin: Bedürfnisse der ZBW-Zielgruppen im Bereich Informationsmanagement. Zusammenfassung der Ergebnisse der qualitativen Stufe. 2010

Tan u. Kumar 2004

TAN, Pang-Ning ; KUMAR, Vipin: Discovery of web robot sessions based on their navigational patterns. In: NING ZHONG, Jiming L. (Hrsg.): Intelligent Technologies for Information Analysis. Springer, 2004, S. 192–222

Wildemuth u. O'Neill 1995

WILDEMUTH, Barbara M. ; O'NEILL, Ann L.: The „known“ in Known-Item Searches: empirical support for user-centered design. In: College & Research Libraries 56 (1995), Nr. 3, S. 265–281

Wilson 2010

WILSON, R. D.: Using clickstream data to enhance business-to-business web site performance. In: Journal of Business & Industrial Marketing 25 (2010), Nr. 3, S. 177–187

Xie u. Joo 2010

XIE, Iris ; JOO, Soohyung: Transitions in Search Tactics During the Web-Based Search Process. In: Journal of the American Society for Information Science and Technology 61 (2010), Nr. 11, S. 2188–2205

Yun 2009

YUN, Gi W.: The unit of analysis and the validity of web log data. In: BERNARD J. JANSEN, Isak T. Amanda Spink S. Amanda Spink (Hrsg.): Handbook of research on web log analysis. Information Science references, 2009, S. 165–180

ZBW 2014

ZBW: Durchsuchte Datenbanken. <https://www.econbiz.de/eb/de/ueber-econbiz/hilfe-datenbanken/>. Version: 2014, Abruf: 2015-03-10