



Universitäts- und Landesbibliothek Düsseldorf

**Elektronische Zeitschriften
in der überregionalen Literaturversorgung**

Teilprojekt ACCELERATE

Abschlussbericht

Inhalt

| | |
|---|----|
| I. Projektauftrag..... | 3 |
| II. Vorbereitung und Durchführung..... | 3 |
| III. Untersuchungsergebnisse..... | 4 |
| IV. Fazit..... | 10 |
| V. Vorträge und Veröffentlichungen..... | 11 |
| VI. Anhang..... | 12 |
| Anhang 1 Entwicklung des Auswertungsalgorithmus'..... | 13 |
| Anhang 2 Die Session-Analyse | 20 |
| Anhang 3 Angebot elektronischer Zeitschriften an der ULB Düsseldorf..... | 25 |
| Anhang 4 Vergleichende Nutzungsuntersuchung von Print- und el. Version | 26 |
| Anhang 5 Nutzung an der ULB Düsseldorf | 28 |
| Anhang 6 Nutzungsbreite des untersuchten Zeitschriftenpaketes..... | 29 |
| Anhang 7 Nutzungsbreite einzelner Titel | 30 |
| Anhang 8 Rechnerbezogene Nutzungsverteilung..... | 32 |
| Anhang 9 Nutzung elektronischer Zeitschriften in versch. Wissenschaftsgebieten | 33 |
| Anhang 10 Nutzungsanteile verschiedener Jahrgänge | 34 |
| Anhang 11 Nutzung auf Artikelebene | 35 |
| Anhang 12 Cross- und Additional Access | 38 |
| Anhang 13 Konsortialnutzung..... | 40 |
| Anhang 14 Grund- und Spitzenversorgung | 41 |

I. Projektauftrag

Das Projekt ACCELERATE (ACCess to ELEctronic liteRATurE) wurde im Zeitraum vom 01.04.1998 bis 12.07.2000 an der Universitäts- und Landesbibliothek Düsseldorf durchgeführt. Es war Teil des übergeordneten Projektes EZUL (Elektronische Zeitschriften in der überregionalen Literaturversorgung), das zum Ziel hatte, Konzepte und Modelle zur Einbindung elektronischer Zeitschriften in die überregionale Literaturversorgung zu entwickeln. Die beiden Projektpartner (Bayerische Staatsbibliothek München und Technische Informationsbibliothek Hannover) befassten sich aus Sicht einer Sondersammelgebietsbibliothek bzw. einer Zentralen Fachbibliothek mit diesem Thema. Das Teilprojekt ACCELERATE dagegen nahm die Perspektive einer Universitätsbibliothek als deren Kunde ein.

Ziel des Düsseldorfer Projektteils war die Untersuchung des Angebots elektronischer Zeitschriften hinsichtlich Akzeptanz, Bereitstellung, Nutzung und Zugriffsformen an einer Universitätsbibliothek als Beitrag zum Entwurf eines Kosten- und Fördermodells zur Integration und Mitfinanzierung elektronischer Zeitschriften durch die DFG im Rahmen der überregionalen Literaturversorgung des SSG-Programms.

II. Vorbereitung und Durchführung

Das Projekt lässt sich grob in drei Phasen untergliedern. Um überhaupt Nutzung und Akzeptanz untersuchen zu können, bedurfte es zunächst eines repräsentativen und konsistenten Angebotes an elektronischen Volltextzeitschriften mit einer entsprechend funktionellen Präsentation. Durch die Mitgliedschaft in mehreren nordrhein-westfälischen Konsortien und durch eigene Vertragsabschlüsse konnte die Universitäts- und Landesbibliothek Düsseldorf schon zu Beginn des Projektes über 2700 elektronische Zeitschriften anbieten. Das Angebot wurde ständig erweitert und bei Projektende konnten unsere Nutzer auf über 3700 Zeitschriften elektronisch zugreifen¹.

Für eine nutzergerechte Präsentation wurden eigene Internetseiten erstellt und in die bewährte Struktur des Düsseldorfer Internetangebotes integriert.

Im zweiten Schritt musste das geschaffene Angebot durch Werbemaßnahmen in der Universität bekannt gemacht werden. Durch verschiedenste Werbemaßnahmen, wie Faltblätter, einen elektronischen Newsletter, Infostand im Foyer, Auslage von Notizzetteln mit entsprechendem Hinweis, Wandplakate, Hinweise an Zeitschriftenauslagen, interne und externe Schulungsangebote, die Einrichtung eines telefonischen Service-Centers u.a.m. wurde auf die neue Dienstleistung aufmerksam gemacht. Ständig steigende Nutzungszahlen² bestätigten den Erfolg der Werbemaßnahmen und schufen die nötige Voraussetzung zur Erstellung aussagefähiger Nutzungsstatistiken.

Gleichzeitig wurden Methoden zur Analyse der Nutzung entwickelt. Aus den Zugriffsprotokollen der Server (LogFiles) musste ein Auswertungsalgorithmus entwickelt werden, der aus den für eine statistische Analyse unbrauchbaren Protokollen die Daten so aufbereitet, dass die Nutzung auf Titel- und Artikelebene herausgefiltert und mit gängigen Datenbank- und Statistikprogrammen ausgewertet werden konnte³.

¹ Siehe Anhang 3: Angebot elektronischer Zeitschriften an der ULB Düsseldorf

² Siehe Anhang 5: Nutzung an der ULB Düsseldorf

³ Siehe Anhang 1: Entwicklung eines Auswertalgorithmus'

Die dritte Phase war durch die Auswertung der erhaltenen Nutzungsdaten und den Vergleich mit alternativen Modellen der Literaturbeschaffung gekennzeichnet. Ein Vergleich der elektronischen Nutzung mit derzeit noch parallel vorhandenen Print-Ausgaben sowie Kosten-/Nutzen-Vergleiche spielten dabei ebenso eine Rolle wie Überlegungen zum Projekt-PEAK der Universität Michigan und des Verlages Elsevier (s. Vorträge und Veröffentlichungen). Außerdem wurde die zentrale Fragestellung nach der Definition von Grund- und Spitzenversorgung bei elektronischen Zeitschriften intensiv diskutiert und daraus ein pay-per-view-Modell zur Einbindung elektronischer Zeitschriften in die überregionale Literaturversorgung vorgeschlagen (siehe Seite 8).

III. Untersuchungsergebnisse

Akzeptanz

Die Frage nach der Akzeptanz ist vielschichtig. Akzeptanz setzt zunächst einmal Kenntnis des Angebotes bei der jeweiligen Nutzergruppe voraus. Zu Beginn des Projektes wurde daher durch verschiedene Werbemaßnahmen intensiv daran gearbeitet, den Bekanntheitsgrad des elektronischen Zeitschriftenangebotes bei der Zielgruppe (Mitarbeiter und Studierende der Heinrich-Heine-Universität Düsseldorf) zu verbessern und elektronische Zeitschriften als neues Dienstleistungsprodukt der ULB Düsseldorf innerhalb der Universität zu etablieren. Stetig steigende Gesamtnutzungszahlen⁴ in dieser Phase dokumentierten den Erfolg dieser Werbemaßnahmen. Detailliertere Untersuchungen der Nutzung zeigten später dann aber große Nutzungsunterschiede in den einzelnen Wissenschaftsdisziplinen, was zum großen Teil sicher durch die stark STM⁵-dominierte Angebotsstruktur bedingt war.⁶ So lagen die Fachbereiche Medizin, Biologie und Chemie deutlich an der Spitze. Darüber hinaus war aber auch zu beobachten, dass vermutlich infrastrukturelle Gegebenheiten das Nutzungsverhalten stark beeinflussen. So fiel auf, dass Institute mit einer großen räumlichen Nähe zum Printbestand in der Nutzung der elektronischen Zeitschriften unverhältnismäßig zurücklagen, während z.B. die stark dislozierte medizinische Fakultät einen hohen Nutzungsgrad verzeichnete, der nicht allein durch die überdurchschnittliche Titellanzahl zu erklären war.

Im Projektantrag noch erwartete technische Einflüsse auf die Akzeptanz wurden durch die fortschreitende Internettechnik gegenstandslos. Akzeptable Zugriffsgeschwindigkeiten, auch von fernen Servern, machen eine lokale Datenhaltung nicht mehr notwendig. Allgemein über das Internet verfügbare Reader ermöglichen die Nutzung aller angebotenen Datenformate. Erwartungsgemäß wurde das derzeit gängigste Format für Volltexte, das pdf-Format, mit 83% aller Volltextzugriffe am meisten verwendet.

Ein Manko war zu Beginn des Projektes allerdings die mangelnde Aktualität des Angebotes mit Verzögerungszeiten der elektronischen Version von bis zu 3 Monaten gegenüber der Printausgabe. Ein Umstand, der besonders Erstnutzer für die weitere Nutzung negativ beeinflussen dürfte. Spätere Untersuchungen zeigten jedoch, dass die Aktualität der angebotenen elektronischen Zeitschriften für die Akzeptanz nur untergeordnete Bedeutung haben dürfte (solange sie nicht deutlich über 3 Monate steigt), denn mit 34% bzw. 44% liegen die Hefte des Vorjahres in der Nutzung deutlich an der Spitze⁷, während Ausgaben des aktuellen

⁴ Siehe Anhang 5: Nutzung an der ULB Düsseldorf

⁵ STM = Science, Technology, Medicine

⁶ Siehe Anhang 3: Angebot elektronischer Zeitschriften an der ULB Düsseldorf

⁷ Siehe Anhang 10: Nutzungsanteile verschiedener Jahrgänge

Jahrgangs, insbesondere in den Geisteswissenschaften deutlich weniger nachgefragt sind. Sicher ein Hinweis darauf, dass Zeitschriftenliteratur häufig gezielt mit einem konkreten Literaturhinweis genutzt wird (der erst einmal veröffentlicht sein muss).

Nutzung

Bei der Untersuchung der Nutzung von elektronischen Zeitschriften ist stets zu fragen, in welchem Umfang die entsprechenden Print-Versionen parallel vorhanden sind. Da bislang noch kaum eine renommierte Zeitschrift ausschließlich online angeboten wird und auch Bibliotheken den Schritt zum reinen Online-Angebot nur selten vollzogen haben, wird das Nutzerverhalten elektronischen Zeitschriften gegenüber quasi immer durch parallele Print-Abonnements beeinflusst. Im Projekt sind beide Medienformen hinsichtlich des Nutzerverhaltens miteinander verglichen worden. In einem Zeitraum von sechs Wochen wurde die Nutzung aller 152 vorhandenen parallelen Print-Titel der mathematisch-naturwissenschaftlichen Fakultät ermittelt, mit Hilfe statistischer Methoden auf ein Jahr hochgerechnet und mit den Nutzungszahlen im elektronischen Bereich verglichen.⁸⁹ Dabei konnte letztlich keine generelle Korrelation zwischen den Nutzungswerten der beiden verschiedenen Medienformen festgestellt werden. Alle Nutzungsverteilungen waren vertreten: Bei beiden Medienformen hohe wie niedrige Nutzung eines Titels, hohe Nutzung der elektronischen Version bei gleichzeitig schwacher Nutzung des parallelen Print-Titels - aber auch umgekehrt.

Als Fazit dieser flankierenden Untersuchung muss daher konstatiert werden, dass das Nutzerverhalten von weit mehr Faktoren beeinflusst wird, als vom reinen Interesse an einem Titel (infrastrukturelle Gegebenheiten, Aufgeschlossenheiten der Institutsmitarbeiter dem neuen Medium gegenüber, Verfügbarkeit der parallelen Print-Publikation u.a.). Eine quantifizierende Wertung dieser Faktoren ist verständlicherweise nicht möglich. Ebenso kann aufgrund der fehlenden Titel-Parität keine grundsätzliche Aussage zur unterschiedlichen Nutzung elektronischer Zeitschriften zwischen Geistes- und Naturwissenschaftlern gemacht werden.

Von besonderem Interesse bei derartigen Nutzungsuntersuchungen ist die Nutzungsverteilung auf die einzelnen Titel. Die Nutzungszahlen 1999 der ULB Düsseldorf, aber auch aller anderen sieben Konsortialbibliotheken zeigen, dass der Großteil des untersuchten Zeitschriftenpaketes (1275 Titel) verhältnismäßig wenig genutzt wurde.¹⁰

⁸ Siehe Anhang 10: Vergleichende Nutzungsuntersuchung von Print- und el. Version

⁹ Siehe auch: Schümmer, Volker:
Nutzungsanalyse von mathematisch-naturwissenschaftlichen Print-Zeitschriften an der
Universitäts- und Landesbibliothek Düsseldorf. Mit einem Ausblick auf die Nutzung
elektronischer Zeitschriften
In: Bibliotheksdienst 33 (1999), H. 9, S. 1475-1493
http://www.dbi-berlin.de/dbi_pub/bd_art/bd_99/99_09_03.htm

¹⁰ Siehe Anhang: 6: Nutzungsbreite des untersuchten Zeitschriftenpaketes und
Anhang 13: Konsortialnutzung

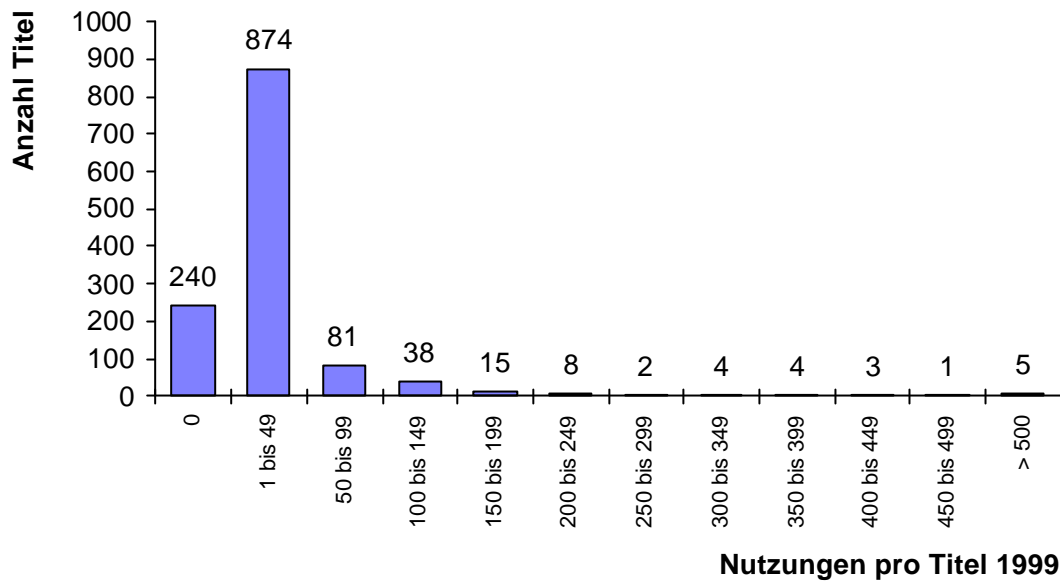


Bild 1: Nutzungsverteilung über das Gesamtangebot von 1275 Zeitschriftentiteln

So wurden 19% aller Titel nicht ein einziges Mal aufgerufen und bei weiteren 68% lag die Nutzung unter 50 Volltextaufrufen pro Jahr, wobei der Großteil auch hier im Bereich zwischen 1 und 9 Nutzungen zu finden ist¹¹. Betrachtet man (durchschnittlich) 1 Nutzung pro Arbeitstag als gut, würden nur 17 Titel (1,3%) diesem Kriterium entsprechen.

Ein Blick auf das Gesamtkonsortium aller acht nordrhein-westfälischen Universitätsbibliotheken bestätigt den hohen Anteil der wenig genutzten Titel: Bei 65,5% lag selbst die konsortiale Gesamtnutzung unter 100 Volltextzugriffen pro Jahr. Dem stehen nur 138 Titel (10,8%) mit einer das o.a. Kriterium erfüllenden Quote von durchschnittlich 1 Nutzung pro Arbeitstag (>300 pro Jahr) gegenüber. 34 Titel wurden gar nicht genutzt¹².

Nutzungsuntersuchungen sprechen in der Regel von einer anonymen Nutzerschaft. Doch gerade unter dem Aspekt der Rentabilität von Zeitschriftenabonnements kommt der Frage "von wievielen verschiedenen Nutzern wird das Angebot genutzt?" eine wichtige Bedeutung zu. Dabei ist die eigentliche Identität des Nutzers unerheblich, wichtiger ist die Nutzungsbreite innerhalb der Universität. Im Rahmen des Projektes wurde daher untersucht, von wieviel verschiedenen Nutzern die meist frequentierten Titel aufgerufen wurden und wie regelmäßig dieser Titel von diesen Nutzern gelesen wird.

Angesichts der Tatsache, dass von der Wissenschaftsseite immer wieder die Wichtigkeit auch weniger bekannter Zeitschriften betont wird, überrascht das Projektergebnis doch ein wenig. Alle Spitzentitel wurden von ca. 60 verschiedenen IP/URL genutzt (wobei davon ausgegangen wird, dass sich hinter jeder IP/URL eine andere Person verbirgt). Ca. zwei Drittel sind allerdings nur mit einer einzigen Nutzung vermerkt; ein Indiz dafür, dass die Nutzung vermutlich zufällig durch ein Retrievalergebnis zustande kam, also nicht gezielt der

¹¹ Siehe Anhang: 6: Nutzungsbreite des untersuchten Zeitschriftenpaketes

¹² Siehe Anhang 13: Konsortialnutzung

Zeitschriftentitel von Interesse war. Lediglich drei bis vier IP/URL ist eine mehr oder weniger regelmäßige Nutzung des Titels zu attestieren¹³.

Obwohl die gemessenen Nutzungswerte, bezogen auf das Gesamtpaket, schon als gering bezeichnet werden müssen, dürfte der Schluss, dass die Projektergebnisse quantitativ auf kostenpflichtige Angebote zu übertragen sind, mit hoher Wahrscheinlichkeit falsch sein, auch wenn man für einen Volltext nur einen geringen Obulus verlangen würde. Die Ergebnisse sind nur in dem Umfeld gültig, in dem die Messungen stattgefunden haben, und das war in diesem Fall durch absolute Kostenfreiheit gekennzeichnet. Es ist davon auszugehen, dass ein nicht unerheblicher Teil der gemessenen Nutzungen eben auf diese Kostenfreiheit zurückzuführen ist, dieser Umstand also Nutzung quasi "erzeugt" hat. Explizite Nutzungsuntersuchungen an kostenpflichtigen Angeboten konnten im Projekt nicht durchgeführt werden, da weder von Seiten der beiden Partnerbibliotheken noch über die Digitale Bibliothek NRW während des Projektzeitraumes derartige Angebote offeriert wurden.

Die in diesem Zusammenhang interessante Differenzierung der Nutzung nach primärem (Mitglieder der Universität) und sekundärem (örtliche Wirtschaft) Nutzungskreis konnte nicht erfolgen, da beim heutigen Stand der Internet-Technik die Herkunft einer Nutzungsanfrage nicht eindeutig zu ermitteln ist. Darüber hinaus sichern sich die Verlage vertraglich gegen Zugriffe von außerhalb des Campus' ab, sodass eine sekundäre Nutzerschaft momentan ohnehin keine Zugriffsmöglichkeit hat.

Grund- und Spitzenversorgung / Fördermodell

Kern des übergeordneten Projektes EZUL war die Entwicklung eines Fördermodells zur Integration elektronischer Zeitschriften in das Programm der überregionalen Literaturversorgung. Gemäß den Förderrichtlinien der DFG soll auch bei dieser neuen Medienform nur die Spitzenversorgung gefördert werden, sodass die Frage zu beantworten war, wie sich Spitzenbedarf in diesem Kontext eigentlich definiert. Im Printbereich gelten solche Titel als Spitzenbedarf, die nur in wenigen Bibliotheken physisch gehalten werden. Dieses Medium bringt es mit sich, dass eine SSG-Bibliothek mit einem (geförderten) Abonnement beliebig viele Dokumentlieferwünsche bedienen kann, ohne dass die Kosten steigen. Der Nutzungsgrad ist also unerheblich. Anders bei elektronischen Zeitschriften: Verlage lassen sich jede Nutzung (Zugriff auf einen einzelnen Artikel) gesondert vergüten, sodass bei diesem Medium eine Definition von Spitzenbedarf eng mit dem Nutzungsaspekt verknüpft ist, eine Untersuchung also von der Nutzungsseite her erfolgen muss. Als Untersuchungsobjekt bot sich ein aus acht nordrhein-westfälischen Universitätsbibliotheken bestehendes Konsortium an, weil die schwache Nutzung eines Titels an einer einzigen Universität zunächst nur als lokales Phänomen zu sehen ist und noch nichts über eine generelle Klassifizierung als Spitzenbedarfstitel aussagt, während eine schwache Nutzung im Gesamtkonsortium aber schon eher ein Indiz in diese Richtung ist. Die Untersuchung der wenig genutzten Titel im Gesamtkonsortium sollte daher zeigen, wie hoch in etwa der Anteil ist, der zum Spitzenbedarf gezählt werden muss, vom Prinzip her also förderungswürdig ist.

¹³ Siehe Anhang 7: Nutzungsbreite einzelner Titel

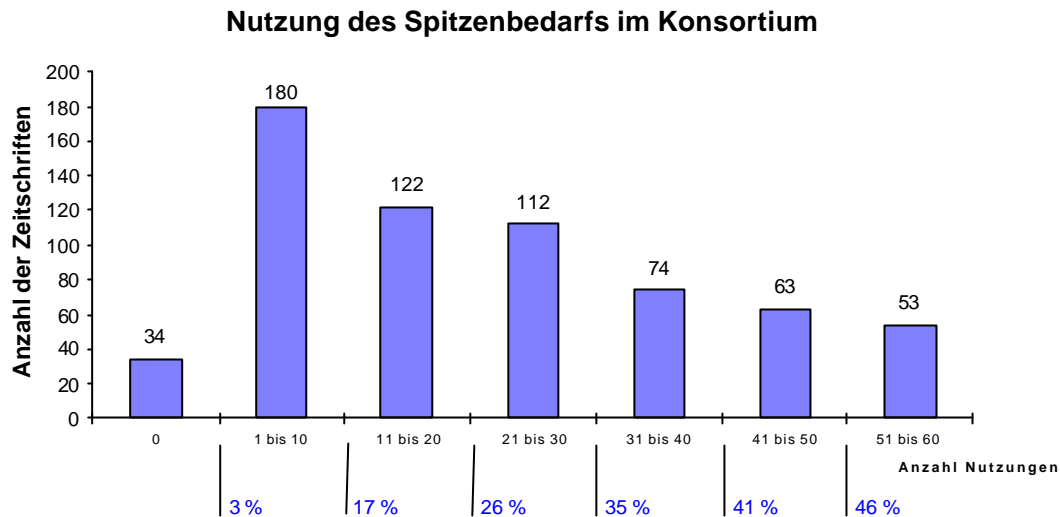


Bild 2: Untersuchung des Spitzenbedarfs im Konsortium

Der absolute Spitzenbedarf muss sicher in den Titeln gesehen werden, die im Beobachtungszeitraum von allen Bibliotheken überhaupt nicht genutzt wurden. Dabei handelt es sich um 3% des Gesamtangebotes. Aber auch Nutzungszahlen im Bereich bis 20 oder 30 Nutzungen pro Jahr sind noch so gering, dass man diese Titel sicher nicht zum allgemeinen Grundbedarf (allenfalls zu einem lokalen Grundbedarf) rechnen kann. Das würde aber bedeuten, dass 26% bzw. 35% aller Titel zum Spitzenbedarf gezählt werden müssten - Werte, die für ein Fördermodell im Rahmen der überregionalen Literaturversorgung inakzeptabel hoch sind. Aufgrund der lokalen Spezifika jeder Universität lassen sich ohnehin keine allgemein gültigen Positivlisten mit Titeln des Spitzenbedarfes erstellen, sodass der Versuch einer analytischen Definition des Spitzenbedarfes nicht zum Erfolg führt. Bei der Entwicklung eines Fördermodells müssten empirisch ermittelte Nutzungen als Kriterium für eine Förderung im Rahmen der überregionalen Literaturversorgung herangezogen werden, will man dem Gedanken der Spitzenbedarfsförderung weiterhin Rechnung tragen.

Untersuchungsmethodik

Die zu Beginn des Projektes eher als nebensächlich betrachtete Frage, wie eine Universitätsbibliothek an statistische Daten über Direktnutzung und –zahlung gelangt, entwickelte sich schon kurz nach Projektstart zum Kernproblem. Ursprünglich als Datenbasis gedachte Nutzungsstatistiken der Verlage waren, wenn überhaupt verfügbar, bei weitem nicht aussagefähig genug. Es gab keine Erhebungen auf Titelebene, sondern nur pauschale, monatliche Zugriffszahlen auf das Gesamtangebot und auch keinerlei Informationen über die Nutzerschaft. Um das Projekt überhaupt durchführen zu können, mussten daher eigene Auswertungsmethoden entwickelt und programmiert werden. Datengrundlage blieben aber sogenannte "LogFiles", in denen sämtliche Zugriffe auf das elektronische Zeitschriftenangebot protokolliert werden. Diese sind grundsätzlich im Besitz desjenigen, der auch physisch die Inhaltsdaten vorhält. In den meisten Fällen sind dies Verlage oder Agenturen und es zeigte sich, dass von deren Seite kein Interesse an einer Weitergabe dieser Rohdaten bestand. So konnte im Projekt nur auf Nutzungsdaten des Elsevier-Verlages zurückgegriffen werden, da dessen elektronische Zeitschriften auf eigenen Servern eines nordrhein-westfälischen Konsortiums angeboten wurden und somit auch die LogFiles verfügbar waren.

Schon der erste im Projekt entwickelte Auswertungsalgorithmus offenbarte eine erhebliche Diskrepanz zu den offiziell von Elsevier erstellten Statistiken. So zeigte sich ein unterschiedliches Verständnis dessen, welche LogFile-Einträge als Nutzung überhaupt gezählt werden sollten und die Definition einer Nutzung wurde zum zentralen Begriff für alle weiteren Untersuchungen. Ferner wurde schnell deutlich, dass, um eine Vergleichbarkeit der Nutzung elektronischer Zeitschriften zu anderen Dokumentlieferformen zu erreichen, die Grundlage eines Auswertungsalgorithmus' die einzelne Session¹⁴ sein muss. Die ACCELERATE-Methode wurde daher in diese Richtung weiterentwickelt mit dem Ergebnis, dass die ermittelten Nutzungszahlen jetzt zwar zum ersten Mal realitätsnah waren, gleichzeitig aber auch stark sanken. Aufgrund der ACCELERATE-Untersuchungen korrigierte Elsevier seinen Auswertungsalgorithmus in einigen Punkten, ohne allerdings die ACCELERATE-Philosophie vollständig zu übernehmen. So blieben beide Statistiken divergent.

Durch diese zweite Weiterentwicklung des Auswertungsalgorithmus' wurde ein Phänomen sichtbar, dass sowohl für Elsevier wie auch für den Springer-Verlag bis dato unbekannt war: nämlich redundante Mehrfacheinträge in den LogFiles, deren Herkunft auch durch intensive Untersuchungen des Springer-Verlages nicht mit Sicherheit geklärt werden konnte. Springer zog daraus den Schluss, dass beim heutigen Stand der Internettechnik zuverlässige Abrechnungssysteme auf der Basis ausschließlich IP-basierter Zugangssysteme nicht möglich sind. Nutzungsstatistiken von elektronischen Zeitschriftenangeboten, die mit der Struktur der Angebote von Elsevier und Springer vergleichbar sind, liegen bei Nichtberücksichtigung dieses Phänomens somit um ca. 30% zu hoch. Der letzte Entwicklungsschritt des ACCELERATE-Algorithmus berücksichtigte auch dieses Phänomen.

Mit diesem Algorithmus wurden dann die in 1999 gesammelten LogFiles ausgewertet. Die so ermittelten Zahlen bildeten die Grundlage für alle Projektergebnisse, insbesondere auch für die Kernfrage, ob und ggf. wie sich Grund- und Spitzenversorgung bei elektronischen Zeitschriften definieren lassen.

Nach Projektende wurde bekannt, dass der Springer-Verlag seine Statistiken über die Nutzung elektronischer Zeitschriften ebenfalls auf die im Projekt entwickelte Technik der Sessionanalyse¹⁵ umstellen will.

¹⁴ Als eine „Session“ soll im Folgenden der Zeitraum bezeichnet werden, in dem sich ein Nutzer ohne längere Unterbrechungen mit dem elektronischen Zeitschriftenangebot beschäftigt.

¹⁵ Siehe Anhang 2: Die Session-Analyse

IV. Fazit

Als Schlussfolgerung aus den vielfältigen Analysen des ACCELERATE-Projektes steht die Erkenntnis, dass Vertragsverhandlungen über einen Konsortialbezug von elektronischen Zeitschriften Nutzungsuntersuchungen als Grundlage haben müssen, um einen optimierten Bestand an Zeitschriften anbieten zu können. Es ist nicht sinnvoll, Gesamtpakete von elektronischen Zeitschriften eines Verlages zu entsprechend hohen Kosten zu beziehen, da der von den Verlagen propagierte Vorteil des Cross- und Additional Access kleiner ist als vielfach angenommen¹⁶. Effizienter dagegen ist ein gezielter Bezug von vielgenutzten Zeitschriften des Grundbedarfs innerhalb eines Konsortiums, um dadurch eine verbesserte Position in Vertragsverhandlungen zu erreichen und Artikel wenig genutzter Zeitschriften über ein pay-per-view-Verfahren zu beschaffen, wobei die Frage einer evtl. Subventionierung hier ausgeklammert bleiben soll.

Es sollte Aufgabe der SSG-Bibliotheken sowie der Zentralen Fachbibliotheken sein, die Zeitschriften, die eine geringe Nutzung aufweisen und somit aus dem Konsortialbezug ausgeklammert werden sollten, im Rahmen der überregionalen Literaturversorgung als Dokumentlieferung bzw. verstärkt als pay-per-view-Angebote bereitzustellen.

Nach abschließender Beurteilung sprechen die Projektergebnisse von ACCELERATE für eine verstärkte Ausrichtung auf pay-per-view-Angebote, wie sie zur Zeit im Gesamtprojekt EZUL aufgebaut und weiterentwickelt werden.

¹⁶ Siehe Anhang 12: Cross- und Additional Access

V. Vorträge und Veröffentlichungen

- *Berg, Heinz-Peter:*
Gewinnung von statistischen Nutzungsdaten aus LogFiles (IuK-Workshop in Regensburg, 5. Februar 1999)
<http://www.bibliothek.uni-regensburg.de/iuk/berg/logfile.htm>
- *Berg, Heinz-Peter, Schäffler, Hildegard, Schröter, Madeleine:*
Elektronische Zeitschriften in der überregionalen Literaturversorgung
In: Bibliotheksdienst 33 (1999), H. 4, S. 608-613
http://www.dbi-berlin.de/dbi_pub/bd_art/bd_99/99_04_05.htm
- *Schümmer, Volker:*
Überlegungen zu Problemen der Grund- und Spitzenversorgung mit gedruckten und elektronischen Zeitschriften
In: Bibliotheksdienst 33 (1999), H. 4, S. 614-628
http://www.dbi-berlin.de/dbi_pub/bd_art/bd_99/99_04_06.htm
- *Karlowitsch, Martin:*
Article-per-view als Alternative zum klassischen Zeitschriftenabonnement?
Überlegungen zum Projekt PEAK
In: Bibliotheksdienst 33 (1999), H. 8, S. 1299-1312
http://www.dbi-berlin.de/dbi_pub/bd_art/bd_99/99_08_06.htm
- *Schümmer, Volker:*
Nutzungsanalyse von mathematisch-naturwissenschaftlichen Print-Zeitschriften an der Universitäts- und Landesbibliothek
Düsseldorf. Mit einem Ausblick auf die Nutzung elektronischer Zeitschriften
In: Bibliotheksdienst 33 (1999), H. 9, S. 1475-1493
http://www.dbi-berlin.de/dbi_pub/bd_art/bd_99/99_09_03.htm
- *Berg, Heinz-Peter, Schäffler, Hildegard, Schröter, Madeleine, Schümmer, Volker:*
Zusammenarbeit von Verlagen und SSG-Bibliotheken im Bereich des überregionalen Zugriffs auf Aufsätze aus elektronischen Zeitschriften: Bericht über einen Workshop - Abdruck und Kommentierung eines Mustervertrages.
In: Bibliotheksdienst 33 (1999), H. 11, S. 1931-1950
http://www.dbi-berlin.de/dbi_pub/bd_art/bd_99/99_11_11.htm
- *Berg, Heinz-Peter:*
Ergebnisse des Projektes ACCELERATE
(Vortrag auf der BDI/DGI-Jahrestagung in Leipzig am 23.03.2000)
<http://www.uni-duesseldorf.de/ulbd/berg/bibtag.htm>
- *Berg, Heinz-Peter:*
"Wird die Nutzung wissenschaftlicher Zeitschriften überschätzt ?"
Ergebnisse des Projektes ACCELERATE
(Vortrag auf dem 1. Symposium des IBLC anlässlich der Frankfurter Buchmesse am 22.10.2000)
<http://www.uni-duesseldorf.de/ulbd/berg/iblc.htm>

VI. Anhang

| | | |
|-----------|---|----|
| Anhang 1 | Entwicklung des Auswertungsalgorithmus'..... | 13 |
| Anhang 2 | Die Session-Analyse | 20 |
| Anhang 3 | Angebot elektronischer Zeitschriften an der ULB Düsseldorf..... | 25 |
| Anhang 4 | Vergleichende Nutzungsuntersuchung von Print- und el. Version | 26 |
| Anhang 5 | Nutzung an der ULB Düsseldorf | 28 |
| Anhang 6 | Nutzungsbreite des untersuchten Zeitschriftenpaketes..... | 29 |
| Anhang 7 | Nutzungsbreite einzelner Titel | 30 |
| Anhang 8 | Rechnerbezogene Nutzungsverteilung..... | 32 |
| Anhang 9 | Nutzung elektronischer Zeitschriften in versch. Wissenschaftsgebieten | 33 |
| Anhang 10 | Nutzungsanteile verschiedener Jahrgänge | 34 |
| Anhang 11 | Nutzung auf Artikelebene | 35 |
| Anhang 12 | Cross- und Additional Access | 38 |
| Anhang 13 | Konsortialnutzung..... | 40 |
| Anhang 14 | Grund- und Spitzenversorgung | 41 |

Entwicklung des Auswertungsalgorithmus'

Grundsätzliches

Ein Großteil der Projektarbeit musste auf die Entwicklung eines geeigneten Auswertungsalgorithmus für LogFiles verwandt werden. Warum war dies nötig angesichts der Tatsache, dass nahezu alle Lizenzverträge über elektronische Zeitschriften eine Verpflichtung des Anbieters zur Lieferung von Statistiken enthalten?

Schon zu Beginn des Projektes wurde deutlich, dass Anbieter ihrer Berichtspflicht in sehr unterschiedlichem Umfang nachkommen. Es gab Statistiken, in denen die Zugriffe bis auf Titelebene ausgewertet waren, was für erste Nutzungsanalysen durchaus ausgereicht hätte. Gerade die unter quantitativen Gesichtspunkten interessantesten Anbieter begnügten sich aber mit Pauschalangaben über die monatlichen Aufrufe des Gesamtangebotes (von mehreren hundert Titeln) bzw. differenzierten nicht einmal zwischen den einzelnen Bibliotheken des betreffenden Konsortiums und boten auch keine Statistik auf Titelebene an. Es zeichnete sich also die Notwendigkeit ab, eigene Auswerteprogramme zu entwickeln, um das Projekt überhaupt durchführen zu können. Dies setzte die Verfügbarkeit von Rohdaten (LogFiles) voraus, die allerdings grundsätzlich immer im Besitz desjenigen sind, von dessen Server die eigentlichen Inhalte angeboten werden; und das sind in der Regel die Anbieter selbst, die, vermutlich aus Wettbewerbsgründen, kein Interesse daran hatten, diese Informationen an Dritte weiterzugeben. Das Projekt konnte schließlich nur deshalb durchgeführt werden, weil aufgrund der mit dem Verlag Elsevier vereinbarten Distributionsform dessen Daten von eigenen Servern eines nordrhein-westfälischen Konsortiums angeboten wurden. Das Konsortium war daher auch im Besitz der LogFiles und stellte sie dankenswerterweise für die Projektarbeit zur Verfügung.

Mit der Überlassung der LogFiles durch das Konsortium war zum einen die Möglichkeit gegeben, gemäß dem Projektauftrag die Nutzung von elektronischen Zeitschriften (beschränkt allerdings auf das Elsevier-Angebot) an der Universitäts- und Landesbibliothek Düsseldorf zu untersuchen, zum anderen enthielten sie aber auch die Zugriffsdaten der anderen sieben Konsortialbibliotheken, was sich als eminent wichtig bei der Untersuchung der Frage nach Grund- und Spitzenbedarf erweisen sollte.

Bis der endgültige Auswertalgorithmus entwickelt war, bedurfte es mehrerer Phasen. Dieses Gebiet derLogFile-Auswertung erwies sich (auch für Verlage) als neu. So gestaltete sich die Entwicklung als iterativer Prozess, in dessen Verlauf insgesamt erst die dritte Version des Algorithmus' zufriedenstellende Ergebnisse zeigte. Die Schritte und Überlegungen, die letztlich zu einer Methode führten, die es ermöglicht, die Nutzung von elektronischen Zeitschriften so realitätsnah wie möglich zu messen, werden im folgenden skizziert.

Alle Auswertungen über Zugriffe im elektronischen Bereich erfolgen mit Hilfe von LogFiles. Das sind Textdateien, die zu jedem Zugriff bzw. Zugriffsversuch auf den WWW-Server verschiedene Informationen protokollieren:

- von wem (welcher IP-Adresse) stammte der Zugriff ?
- wann erfolgte der Zugriff ?
- auf welches File wurde zugegriffen (request) ?
- wie hat der Server auf die Anfrage reagiert ?

- wie groß war das Antwort-File ?

Diese Struktur ist als *Common LogFile Format (CLF)* genormt und wird von den meisten WWW-Servern verwendet.

1. Entwicklungsstufe

Diese Informationstiefe reicht für allgemeine statistische Auswertungen, wie sie von bekannten Statistikprogrammen wie Analog oder WebAlyzer produziert werden, aus. Für die im Projekt durchzuführenden Untersuchungen war sie allerdings nicht hinreichend, weil die eigentliche Information, auf welchen Titel/Artikel zugegriffen wurde, noch innerhalb des request verschlüsselt ist. Die Aufgabe war also zunächst, den request, der bei jedem Anbieter anders aussieht, zu entschlüsseln, um die Nutzdaten von den irrelevanten Einträgen trennen zu können. Aufgrund der immer größer werdenden LogFile-Größen (bis 270 MB) bot es sich an, zunächst alle Einträge herauszufiltern, die keine positive Server-Antwort enthielten.

Anm.: In einem LogFile werden (je nach Einstellung des WWW-Servers) nicht nur positiv erledigte Anfragen protokolliert, sondern auch alle Anfragen, auf die der Server aus den verschiedensten Gründen mit einer Fehlermeldung reagierte (also z.B. wenn dem Nutzer die benötigte Zugriffsberechtigung fehlte). Erkennbar ist dies anhand des HTTP-Status-Codes, in dem die Antwort des Servers aus eine Anfrage verschlüsselt abgelegt ist.

Positive Server-Antworten erhalten den HTTP-Status-Code 200 (OK) bzw. 304 (Antwort durch Proxy-Server). Durch das Herausfiltern der LogFile-Einträge mit diesen beiden Status-Codes fielen 40-70% aller Einträge von vornherein heraus, was zeigt, in welchem hohem Maß von nicht autorisierten Benutzern versucht wurde, auf die elektronischen Elsevier- Zeitschriften in NRW zuzugreifen. Überhaupt ist ein Blick auf die Zusammensetzung eines LogFiles von Elsevier interessant:

- ca. 40% für Auswertezwecke relevante LogFile-Einträge (alle Zugriffe, die von Seiten des Servers nicht mit einer Fehlermeldung quittiert wurden); davon aber nur ca. 3% aus dem Konsortium, das Zugriff auf die Volltexte hatte. Der Rest sind zum großen Teil (frei zugängliche) Zugriffe auf Abstracts und Inhaltsverzeichnisse.
- ca. 15% Image-Files (die Übermittlung von auf einzelnen Seiten integrierter kleiner Bildchen, z.B. der Titelseite eines Heftes im Kleinformat).
- ca. 45% für Auswertezwecke irrelevante Einträge (mit einer Fehlermeldung quittierte Zugriffsversuche). Diese hohe Zahl kam vermutlich dadurch zustande, dass Suchmaschinen von außerhalb des Konsortiums häufig versuchten, systematisch Volltexte herunterzuladen.

Nach dieser ersten Filterung wurden die Nutzdaten den einzelnen Konsortialbibliotheken zugeordnet. Die IP/URL jedes LogFile-Eintrages wurde dazu mit einer Tabelle verglichen, die neben den einzelnen Campus-Sites (z.B. uni-duesseldorf.de) auch alle IP's enthielt, die berechtigt waren, auf das Volltextangebot zuzugreifen.

Der nächste Schritt bestand in der Auswertung des request, um die einzelnen Transaktionen (z.B. Volltextzugriff, Abstractzugriff, Suchanfrage etc.) voneinander unterscheiden zu können. Da über die Struktur keinerlei Informationen vorlagen, mussten zunächst die verschiedenen request-Muster erkannt und decodiert werden. Dazu wurde jedes einzelne als Adresse im WWW-Browser eingetragen, um festzustellen, zu welcher Transaktion es gehört. So konnten 24 verschiedene Muster von LogFile-Einträgen ermittelt werden, die zum Teil eine bestimmte Transaktion wiedergaben, zum Teil aber auch nur aus technischen Gründen

gesondert vermerkt wurden (auch wenn sie immer mit einer Transaktion in Zusammenhang stehen):

| Abkürzung | Transaktion |
|------------------|---|
| ABST | Abstractzugriff |
| ACIT | Bibliographische Daten eines Artikels |
| AGIF | Eingebettete Graphik bei der Übertragung einer Seite im tiff-Format |
| ALPH | Alphabetische Titelliste |
| APDF | Artikelzugriff im pdf-Format |
| APG | Zugriff auf eine Artikelseite im tiff-Format |
| APGL | Wie APG, nur vergrößerte Darstellung |
| APS | Artikelzugriff im postscript-Format |
| ATH | Darstellung von Artikelseiten als kleine Bildchen (nicht lesbar) |
| DBIN | Database Info |
| HELP | Zugriff auf eine Hilfeseite |
| IGIF | Wie AGIF, aber bezogen auf ein ganzes Heft |
| IPDF | Download eines ganzen Heftes im pdf-Format |
| IPG | Zugriff auf eine Artikelseite bei Download eines ganzen Heftes |
| IPGL | Wie IPG, nur vergrößerte Darstellung |
| IPS | Download eines ganzen Heftes im postscript-Format |
| ITH | Wie ATH, aber bezogen auf ein ganzes Heft |
| MAIN | Aufrufen der Hauptseite |
| SEXP | Expanded Search |
| SRES | Search Result |
| SSIM | Simple Search |
| SYST | Systematische Titelliste |
| TOC | Inhaltsverzeichnis eines Heftes |
| VIL | Hauptseite einer Zeitschrift |

Anm.: Das Ergebnis dieser Untersuchung ist nicht auf LogFiles anderer Anbieter übertragbar, da deren Dateistrukturen auf dem Server völlig anders aussehen und sich die Bezeichnungen der einzelnen Dokumente selbstverständlich auch unterscheiden. Und selbst auf andere Elsevier-Server ist dieser Auswertalgorithmus nur dann anwendbar, wenn dieselbe Version der Retrieval-Software verwendet wird. Sollte sich auch auf dem NRW-Server die Retrieval-Software nur geringfügig ändern, würde das wahrscheinlich die Unbrauchbarkeit dieses Teils des Algorithmus' zur Folge haben.

Soweit aufgeschlüsselt war es nun möglich, alle, einen Volltextzugriffe repräsentierenden LogFile-Einträge herauszufiltern und anhand einer ISSN-Tabelle zu einem Titel-Ranking zu addieren. Damit war die erste Entwicklungsstufe abgeschlossen, die (durch einfache Addition der die Volltextzugriffe repräsentierenden LogFile-Einträge) dasselbe Prinzip anwendete, die auch den meisten Verlagsstatistiken zugrunde liegen (Additionsmethode).

Doch schon dieser erste Algorithmus lieferte Zahlen, die erheblich unter den offiziell von Elsevier vorgelegten Statistiken lagen. Als Grund zeigte sich ein unterschiedliches Verständnis dessen, welche LogFile-Einträge als Volltextzugriffe zu zählen waren. So ist die Vorgehensweise bei Verlagsstatistiken in der Regel so, dass alle LogFile-Einträge, die in irgendeiner Form eine Volltextnutzung repräsentieren, einfach addiert werden - im günstigsten Fall noch differenziert nach Titel, häufig aber auch einfach als Summe, bezogen auf das gesamte Verlagspaket. Dabei bleibt unberücksichtigt, ob sich der Benutzer nur eine Seite eines Artikels, einen ganzen Artikel oder gar ein ganzes Heft heruntergeladen hat. Das führt z.B. dazu, dass bei einem Benutzer, der einen 12-seitigen Artikel Seite für Seite herunterlädt (bei der Anzeige im tiff-Format nur so möglich), 12 Nutzungen gezählt werden, während bei einem anderen, der denselben Artikel im pdf-Format aufruft, nur eine einzige Nutzung nachgewiesen wird. Im Gegensatz zu den offiziellen Verlagsstatistiken enthielt aber schon

dieser erste Accelerate-Algorithmus Korrekturmechanismen, die zumindest einen Teil dieser Zählproblematik berücksichtigten.

Schon in dieser Phase zeigte sich, dass der Begriff der Nutzung zum zentralen Definitionsproblem werden würde. Nicht zu unrecht wiesen Verlagsvertreter daraufhin, dass sich Nutzung nicht ausschließlich auf Volltextzugriffe beschränke, sondern dass selbstverständlich auch das Lesen von Abstracts und Inhaltsverzeichnissen eine Nutzung darstelle, ja selbst das Browsing, vergleichbar etwa mit dem Blättern in Print-Zeitschriften, als Nutzung gesehen werden müsse; eine Statistik ausschließlich der Volltextzugriffe demnach zu kurz greife. Dem musste aus Projektsicht entgegengehalten werden, dass diese unterschiedlichen Nutzungsformen erstens nicht in einer einzigen Statistik zusammengefasst werden können, zweitens gegeneinander nicht bewertet werden können und drittens auch die Verlage selbst nur Volltextzugriffe statistisch auswerten, vermutlich, weil nur diese - als einzig kostenpflichtige - Nutzungsform für eine Vermarktung interessant ist. Alle Untersuchungen basieren aus diesen Gründen ausschließlich auf Volltextzugriffen.

2. Entwicklungsstufe

Nutzungsstatistiken für elektronische Zeitschriften haben zwar, absolut gesehen, schon einen Aussagewert, sie erhalten aber erst dann eine für die Erwerbungspolitik relevante Bedeutung, wenn ihre Aussagen mit denen anderer Dokumentliefersysteme vergleichbar werden und es möglich wird, verschiedene Dienste in ein Erwerbungsmodell zu integrieren. Diese Vergleichbarkeit erfordert aber zunächst eine gemeinsame Basis. Die schon im Bereich der Dokumentlieferung verwendete "Einheit" der Nutzung eines einzelnen Artikels bietet sich hier an. Gefordert werden muss also ein Auswertalgorithmus, der auf der Basis dieser Einheit zählt. Herkömmliche Zählverfahren leisten dies nicht, da sie nicht berücksichtigen, "wer" auf einen Artikel zugegriffen hat. So ist es z.B. ein Unterschied, ob ein Artikel von drei verschiedenen Nutzern aufgerufen wird oder von einem einzigen (innerhalb kurzer Zeit) dreimal. Im zweiten Fall dürfte nur eine Nutzung gezählt werden, da der Benutzer beim zweiten und dritten Aufruf keine neue Information mehr erhalten hat. Oder anders ausgedrückt: Pro Dokumentlieferung (oder wenn der Volltextzugriff kostenpflichtig wäre) hätte der Benutzer sich den Artikel auch nur einmal bestellt.

Das bedeutete für die Entwicklung des Auswertalgorithmus', dass der Begriff der "Session" zentrale Bedeutung erlangte. Alle Volltextzugriffe mussten innerhalb ihres Kontextes, der Session, gesehen werden und waren nur dann als Nutzung zu zählen, wenn sie eine für den Nutzer neue Information repräsentierten. Ziel war es daher, solche Artikelaufrufe nicht weiter als eigenständige Nutzung zu zählen,

- die innerhalb einer Session mehrfach erfolgten.
- die innerhalb einer Session wiederholt, aber in einem anderen Datenformat erfolgten.
- bei denen innerhalb einer Session jede Seite einzeln aufgerufen wurde. In diesem Fall wurde nur **eine** Nutzung, bezogen auf den Artikel gezählt.

Die Schwierigkeit bei der Umsetzung bestand zunächst darin, eine Session genau zu umreißen, d.h. alle LogFile-Einträge zu bestimmen, die zu einer Session gehörten. Da es eine eindeutige, unangreifbare Lösung dieses Problems nicht gibt, wurde ein pragmatischer Ansatz gewählt: Ausgehend von der Überlegung, dass es im DV-Bereich vielfach üblich ist, einen Benutzer, der eine bestimmte Zeit lang keine Eingabe getätigt hat, automatisch auszuloggen

(also seine Session zu beenden), wurde versucht, einen solchen, geeigneten Zeitraum anhand der LogFiles zu bestimmen. Es zeigte sich, dass eine Session mit großer Wahrscheinlichkeit beendet war, wenn 19 Minuten lang keine Eingabe mehr erfolgte. Die DV-technische Umsetzung erfolgte dann in folgenden Schritten:

1. Alle Einträge im LogFile wurde nach URL/IP (1.Kriterium) und Zugriffszeit (2. Kriterium) sortiert.
Ergebnis: eine nach IP sortierte, chronologische Liste aller LogFile-Einträge.
2. Anhand dieser Liste wurde jedem LogFile-Eintrag eine laufende Sessionnummer zugeordnet. Dabei wurde diese Nummer immer dann um 1 inkrementiert, wenn sich beim nächsten LogFile-Eintrag entweder die URL/IP änderte oder die Zeit zwischen zwei Transaktionen 19 Minuten überschritt.
Ergebnis: eine über eine Datenbank darstellbare Aufstellung aller Sessions mit ihren zugehörigen Transaktionen.

Anm.: Fehlerbetrachtung:

Eine auf diese Weise erfolgte Ermittlung der Sessions kann aus folgenden Gründen zu Fehlschlüssen führen:

1. Der Benutzer wechselt innerhalb der 19 Minuten, ohne dass es im LogFile erkennbar wird (z.B. wenn die Benutzung des PC's innerhalb eines PC-Pools wechselt).
In diesem Fall würden zwei (oder mehrere) Benutzer fälschlicherweise in einer Session zusammengefasst. Auswirkungen auf die Statistik würden aber nur auftreten, wenn der zweite (oder weitere) Benutzer zufällig genau einen der Artikel aufruft, die schon sein Vorgänger aufgerufen hat. In diesem Fall würde die Nutzung nicht gezählt. In Anbetracht der Tatsache, dass insgesamt 81% aller aufgerufenen Artikel nur einmal aufgerufen wurden, ist dieses Szenario extrem unwahrscheinlich. Und sollte dieser seltene Fall tatsächlich einmal auftreten, hätte dieser Messfehler keine signifikante Auswirkung auf die Statistik.
2. Der Benutzer tätigt tatsächlich erst nach mehr als 19 Minuten die nächste Eingabe (z.B. nachdem er einen längeren Artikel am Bildschirm gelesen hat).
In diesem Fall würde eine Benutzung fälschlicherweise auf zwei Sessions aufgeteilt. Auswirkungen auf die Statistik würden nur auftreten, wenn der Benutzer noch einmal einen der Artikel aufruft, die er schon aufgerufen hatte. Nur dann würde eine Nutzung zuviel gezählt. Ein auch nicht sehr wahrscheinlicher Fall, der ebenfalls keine signifikante Auswirkung auf die Statistik hätte.

Obwohl dieses Verfahren also nicht 100%-ig genau ist, zeigt die Fehlerbetrachtung seine Zulässigkeit.

Anm.: Für Abrechnungszwecke ist die erzielbare Genauigkeit nicht hinreichend. Der Springer-Verlag hat daraus den Schluss gezogen, dass ein auf LogFiles basierendes Abrechnungssystem nur bei individueller Anmeldung realisierbar ist.

3. Entwicklungsstufe

Mit der Integration der Sessionanalyse in den Auswertalgorithmus war das wichtigste Element des Auswertalgorithmus' entwickelt und es wurde möglich, statistische Nutzungszahlen zu erzeugen, die einen realistischen Blick auf den Informationsbedarf zulassen und durch dessen Definition von Nutzung auch eine Vergleichbarkeit mit anderen Dokumentlieferdiensten möglich wird. Durch die chronologische Darstellung einzelner Sessions wurde aber auch ein bis dahin unbekanntes Phänomen sichtbar: das Auftreten redundanter Mehrfacheinträge. Die folgende Abbildung zeigt ein prägnantes Beispiel:

Sessionanalyse Springer-LINK

Sessionstatistik:

Sessionnummer: 560 APDF: 6 VIL: 0 ALPH: 0
 Einträge: 10 ABST: 3 SSIM: 1 HELP: 0
 aufgerufene Titel: 3 TOC: 0 INF: 0 FORUM: 0
 Sessiondauer: 00:13:07 INDX: 0 SYST: 0 SONST: 0

Volltextstatistik:

verschiedene Zeitschriften: 2
 verschiedene Hefte: 2
 verschiedene Artikel: 2

| IP | LogDatum | Titel | Issue | Artikel | Kategorie | HTTP-Status | Bytes | |
|-----|---------------|-------------------|-------|---------|-----------|-------------|-------|--------|
| 560 | 134.99.254.23 | 28.12.98 21:42:12 | no | no | no | SSIM | 200 | 3146 |
| 560 | 134.99.254.23 | 28.12.98 21:45:14 | 00117 | 8038007 | 80380554 | ABST | 200 | 5168 |
| 560 | 134.99.254.23 | 28.12.98 21:45:23 | 00117 | 8038007 | 80380554 | APDF | 200 | 473392 |
| 560 | 134.99.254.23 | 28.12.98 21:45:25 | 00117 | 8038007 | 80380554 | APDF | 200 | 473392 |
| 560 | 134.99.254.23 | 28.12.98 21:48:16 | 00117 | 8038007 | 80380554 | APDF | 200 | 473392 |
| 560 | 134.99.254.23 | 28.12.98 21:50:49 | 00234 | 5037003 | 50370219 | ABST | 200 | 4519 |
| 560 | 134.99.254.23 | 28.12.98 21:51:06 | 00106 | 6044012 | 60440666 | ABST | 200 | 6315 |
| 560 | 134.99.254.23 | 28.12.98 21:51:16 | 00106 | 6044012 | 60440666 | APDF | 200 | 900255 |
| 560 | 134.99.254.23 | 28.12.98 21:51:18 | 00106 | 6044012 | 60440666 | APDF | 200 | 900255 |
| 560 | 134.99.254.23 | 28.12.98 21:55:19 | 00106 | 6044012 | 60440666 | APDF | 200 | 900255 |
| 560 | | | | | | | | |

Datensatz: 560 von 634

Formularansicht

Bild 3: Beispiel für redundante Mehrfacheinträge

Erläuterung:

In dieser Session wurden zwei verschiedene Artikel im pdf-Format aufgerufen, allerdings jeweils dreimal. Es wird sichtbar, dass identische Volltextartikel mehrfach innerhalb einer Session heruntergeladen wurden. Oftmals lagen die einzelnen Downloads nur wenige Sekunden auseinander, sodass zunächst eine technische Ursache vermutet wurde. Umfangreiche Testreihen des Springer-Verlages bestätigten diese Vermutung allerdings nicht, sodass als einzige Erklärung blieb, dass der Nutzer selbst wiederholt einen Artikel angeklickt hatte (vielleicht weil ihm die erste Übertragung zu lange dauerte). Abstände von einer Sekunde wurden durch einen (unnötigen) Doppelklick erklärt.

Wenn diese Erklärung richtig war, musste dieses Phänomen auch bei anderen Anbietern mit ähnlicher Angebotsstruktur zu beobachten sein, denn das Nutzerverhalten ist anbieterunabhängig. Untersuchungen der Elsevier-LogFiles zeigten dann auch tatsächlich dasselbe Phänomen mit der Konsequenz, dass eine dritte Stufe des Auswertalgorithmus' entwickelt werden musste, die redundante Mehrfacheinträge nur noch einmal zählt.

Zusammenfassung

Mit Realisierung der dritten Entwicklungsstufe war ein Auswertalgorithmus entstanden, der im Gegensatz zur herkömmlichen Additionsmethode drei wesentliche Verbesserungen enthielt:

- Berücksichtigung, ob ein Artikel in einzelnen Seiten oder komplett aufgerufen wurde.
- Berücksichtigung, ob derselbe Artikel in verschiedenen Datenformaten aufgerufen wurde.
- Berücksichtigung redundanter Mehrfacheinträge.

Folge dieses aufwendigen Verfahrens war allerdings, dass die gemessenen Nutzungszahlen um 30-50% unter den offiziellen Nutzungsstatistiken der Anbieter lagen, die deswegen nicht falsch waren, aber eben nur die undifferenzierte Anzahl der vom Server gelieferten Volltextdokumente ausweisen, was dem allgemeinen Verständnis von Nutzung nicht gerecht wird.

Es stellte sich abschließend die Frage nach der Ursache dieser unterschiedlichen Messergebnisse.

Relativ genau zu bestimmen war der "Messfehler", der durch redundante Mehrfacheinträge entsteht. Dazu mussten nur beide Ergebnisse, die aus den Springer-LogFiles gewonnen werden konnten, gegenübergestellt werden, da der Springer-Verlag zur damaligen Zeit ausschließlich Volltexte im pdf-Format anbot und somit keine Überlappungen mit den beiden anderen "Verbesserungen" auftreten konnten. Es zeigte sich eine erstaunliche monatliche Konstanz dieses "Messfehlers":

| | |
|------------------|--------------|
| Dez 1998 | 71,3% |
| Jan 1999 | 69,1% |
| Jun 1999 | 72,7% |
| Jul 1999 | 76,0% |
| Aug 1999 | 70,0% |
| | |
| Im Mittel | 71,8% |

Aus der Konstanz dieser Stichproben ist nicht nur eine Verallgemeinerung auf die Nutzung des Springer-Angebotes ableitbar, es lässt auch vermuten, dass es sich als determiniertes Phänomen auf Angebote anderer Anbieter, die elektronische Zeitschriften in ähnlicher Form anbieten, übertragen lässt.

So lassen sich die erheblichen Diskrepanzen zwischen offizieller Elsevier-Statistik (=100%) und Accelerate-Algorithmus (ca. 50%) wie folgt begründen:

- Berücksichtigung redundanter Mehrfacheinträge: minus 30%.
- Berücksichtigung verschiedener Dateiformate und Textumfänge: minus 20%.

Die Session-Analyse

Idee und Entwicklung

Kernstück des im Projekt entwickelten Auswertungsalgorithmus' ist die sogenannte Session-Analyse. Dahinter steht die Erkenntnis, dass aussagefähige Nutzungsstatistiken nur zu erhalten sind, wenn man das Nutzungsverhalten des einzelnen Nutzers analysiert und dadurch berücksichtigen kann, ob der einzelne Volltextaufruf dem Nutzer wirklich einen Informationsgewinn verschafft hat oder ob dieser Nutzer den angeforderten Text mehrfach erhalten hat.

Als Hilfsmittel wurden die einzelnen Sessions graphisch dargestellt. Dazu erzeugte der Auswertungsalgorithmus aus den LogFiles mehrere Textfiles mit statistischen Informationen. Das Format für die einzelnen Datenfelder wurde so gewählt, dass sich die Textfiles in ein Datenbankprogramm importieren lassen (MS Access), mit dessen Hilfe statistische Auswertungen erst möglich werden.

Das Ergebnis ist eine chronologische Darstellung jeder einzelnen Session:

Sessionanalyse Elsevier

Sessionstatistik:

Sessionnr.: **bapr98197** TOC: 2 APDF: 2 IPDF: 0
 Einträge: **9** ABST: 2 APS: 0 IPS: 0
 Sessiondauer (in s): **190** APG: 0 IPG: 0
 APGL: 0 IPGL: 0

Statistik für Vergleich mit Dokumentlieferung:

Verschiedene Artikel: 2
 Verschiedene Hefte: 2
 Verschiedene Zeitschriften: 1

Statistik für Vergleich mit Print-Nutzung:

Verschiedene Hefte: 2
 Verschiedene Zeitschriften: 1

| IP | LogDatum | Titel | Issue | Artikel | Kürzel | Page | Status | Bytes |
|--------------------------|-------------------|----------|----------|----------|--------|------|--------|---------|
| pc.ub.uni-duesseldorf.de | 15.04.98 19:35:58 | no | no | no | MAIN | no | 200 | 2845 |
| pc.ub.uni-duesseldorf.de | 15.04.98 19:36:09 | no | no | no | SYST | no | 200 | 17295 |
| pc.ub.uni-duesseldorf.de | 15.04.98 19:36:18 | 87563282 | no | no | VIL | no | 200 | 5451 |
| pc.ub.uni-duesseldorf.de | 15.04.98 19:36:24 | 87563282 | v0022i02 | no | TOC | no | 200 | 9116 |
| pc.ub.uni-duesseldorf.de | 15.04.98 19:36:33 | 87563282 | v0022i02 | 97002342 | ABST | no | 200 | 7304 |
| pc.ub.uni-duesseldorf.de | 15.04.98 19:37:35 | 87563282 | v0022i02 | 97002342 | APDF | 0 | 200 | 1426040 |
| pc.ub.uni-duesseldorf.de | 15.04.98 19:38:23 | 87563282 | sz978592 | no | TOC | no | 200 | 8483 |
| pc.ub.uni-duesseldorf.de | 15.04.98 19:38:24 | 87563282 | sz978592 | 97001270 | ABST | no | 200 | 6330 |
| pc.ub.uni-duesseldorf.de | 15.04.98 19:39:08 | 87563282 | sz978592 | 97001270 | APDF | 0 | 200 | 1245214 |

Datensatz: 110 von 15184

Bild 4: Chronologische Darstellung einer einzelnen Session

Erläuterung:

Zu sehen ist der chronologische Ablauf einer Session vom 15.04.1998, die über einen PC aus dem Informationszentrum der ULB Düsseldorf durchgeführt wurde. Die insgesamt 9 Einträge geben die einzelnen Transaktionen der Session wieder:

1. MAIN Einstieg über die Hauptseite des Elsevier-Angebotes
2. SYST Systematische Titelliste
3. VIL Hauptseite der Zeitschrift mit der ISSN 87563282 (Bone)
4. TOC Inhaltsverzeichnis des Heftes mit der Nummer v0022i02
5. ABST Lesen des Abstracts mit der Nummer 97002342
6. APDF Aufruf dieses Artikels im Volltext
7. TOC Aufruf eines anderen Inhaltsverzeichnisses
8. ABST Lesen eines weiteren Abstracts
9. APDF Wiederum Aufruf des zugehörigen Volltextartikels

So entstand für jede Session ein chronologischer Ablauf aller darin durchgeführten Transaktionen, anhand derer das konkrete Nutzungsverhalten studiert werden konnte (z.B. in wievielen Sessions gab es überhaupt Volltextzugriffe; lag dem Volltextzugriff vermutlich eine konkrete Quelle zugrunde oder war die Auswahl dieser bestimmten Zeitschrift zufällig durch ein Rechercheergebnis bedingt u.a.). So wurde in über der Hälfte der Sessions (54 %) mindestens ein Volltext aufgerufen, während in 46% aller Fälle das elektronische Zeitschriftenangebot ohne Volltextaufruf genutzt wurde. Dass im Internet vieles ausprobiert und "nur mal eben angeklickt" wird, war daran zu belegen, dass in 27% aller Sessions weder Volltextaufrufe noch Tables of Contents noch Abstract-Aufrufe beinhalteten.

Nachfolgende Grafik zeigt, wieviele Volltextnutzungen pro Session erfolgten:

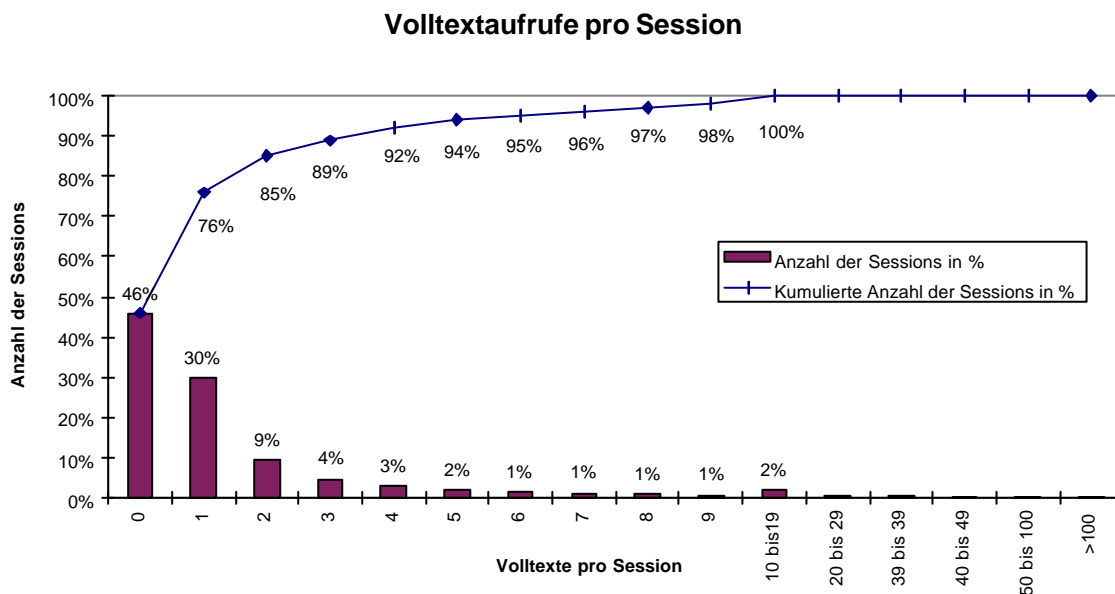


Bild 5: Häufigkeitsverteilung der Volltexte pro Session

Eine Session dauerte durchschnittlich 8 Minuten und 22 Sekunden und es wurden im Mittel 1,69 Volltexte pro Session aufgerufen.

Anteil der Volltextnutzung an der Gesamtnutzung

Die Untersuchung der Frage, in wievielen Sessions überhaupt Volltexte aufgerufen wurden, sollte das Nutzerverhalten qualitativ beleuchten. Festgestellt wurde, dass knapp die Hälfte aller Sessions ohne Volltextnutzung stattfand. Ein starkes Indiz dafür, dass der Browsing-Anteil als eine Nutzungsform sehr hoch sein dürfte. Ca. ein Drittel aller Sessions wiesen genau eine Volltextnutzung auf. Hier ist zu vermuten, dass der Nutzung entweder eine konkrete Quellenangabe zugrunde lag oder sie als Rechercheergebnis zustande kam. Und 24 % aller Sessions enthielten mehr als eine Volltextnutzung. Hier sind vermehrt manuelle oder automatische Themensuchen anzutreffen bis hin zu automatischen Suchrobotern, die grundsätzlich jeden Artikel einer Zeitschrift oder mit einem bestimmten Stichwort herunterladen.

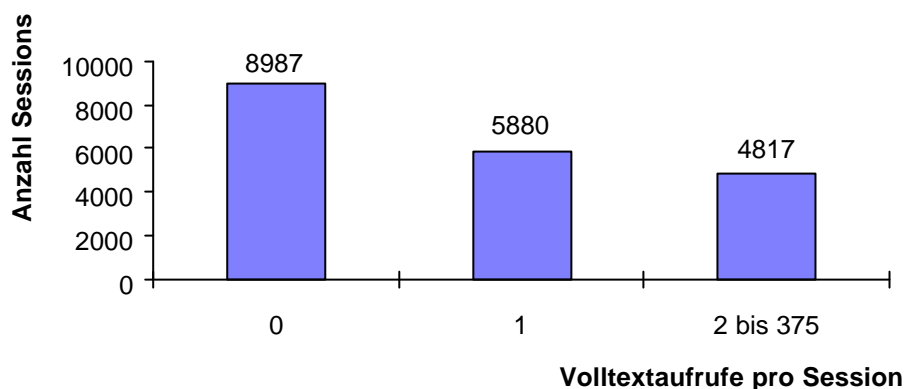


Bild 6: Anzahl der Volltextaufrufe, die pro Session erfolgten

Nutzerverhalten

Im Mittelpunkt dieser Untersuchung stand die Art und Weise, wie Nutzer mit dem elektronischen Zeitschriftenangebot umgehen bzw. der Versuch herauszufinden, mit welcher Intention das Angebot genutzt wird. Obwohl diese Untersuchung spekulativen Charakter hat, lassen sich anhand der chronologisch dargestellten Transaktionen jeder Session doch verschiedene Nutzungsformen unterscheiden:

1. Heft sichten:
Um auf dem aktuellem Stand der Forschung zu bleiben, werden aktuelle Hefte auf relevante Artikel durchgesehen. Auch Sessions ohne Volltextaufrufe (also z.B. nur Inhaltsverzeichnisse oder Abstracts) spielen hier ein Rolle.
Für diese Kategorie wurden die Hefte des Jahrganges 1999 als aktuelle Hefte definiert, allerdings wird eine Zeitverzögerung von drei Monaten berücksichtigt. Findet die Session z.B. im Januar bis März 1999 statt, werden Aufrufe von 98-er-Heften (Oktober bis Dezember) noch als aktuell gewertet.
2. Suchergebnis:
Volltextaufruf, der aufgrund einer Suchanfrage zustande kam.

3. Browsing:
Hier handelt es sich um unspezifisches Anklicken von Zeitschriften, Tables of Contents, Abstract Volltextformaten ohne erkennbare Struktur.
4. Suchmaschine:
Das regelmäßige, strukturierte Laden vieler oder aller Artikel eines Heftes bzw. einer oder mehrerer Zeitschriften in kurzer Zeit wurde als maschineller Zugriff gewertet.
5. Themasuuche:
In einer oder mehreren Zeitschriften wurden die Inhaltsangaben nach relevanten Artikeln durchsucht, teils werden Artikel geladen. Die Hefte sind allerdings nicht so aktuell wie bei „Hefte sichten“.
6. DocDel:
Diese Kategorie ist einer Dokumentlieferung vergleichbar und wurde dann vergeben, wenn die Session ein zielgerichtetes Vorgehen auf einen bestimmten Artikel hin zeigte. In diesem Fall wurde davon ausgegangen, dass eine konkrete Quellenangabe vorlag.
7. unklar:
Hier wurden alle Sessions eingeordnet, die in keine andere Kategorie passten.

Diese Auswertung erfolgte manuell, indem eine Session zufällig ausgewählt und kategorisiert wurde. Um ein repräsentatives Ergebnis zu erhalten, wurden über 4100 Session (ca. 21%) berücksichtigt. Folgende Grafik zeigt das Ergebnis:

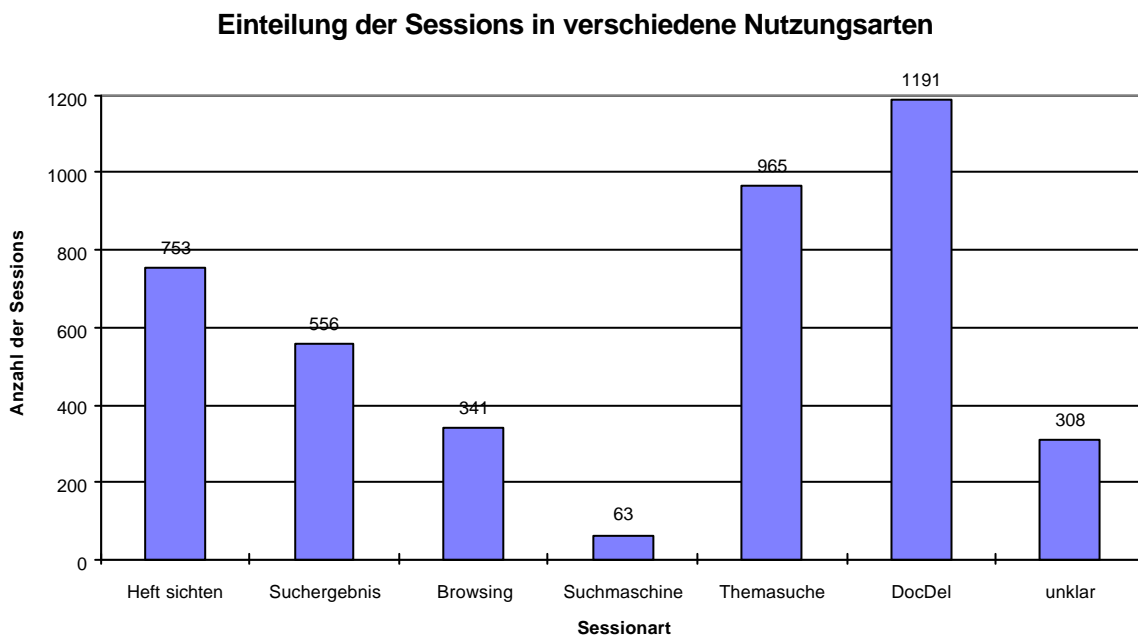


Bild 7: Verteilung der verschiedenen Nutzungsarten

29% der so kategorisierten Sessions zeigten eine Struktur, die vermuten lässt, dass der Nutzung eine konkrete Quellenangabe zugrunde lag. Bei 23% war zu vermuten, dass Hefte

auf ein bestimmtes Thema hin durchsucht wurden und 18% aller Nutzer interessierten sich hauptsächlich für die neueste Literatur (vermutlich in ihrem Fachgebiet). 13% sahen sich einen Volltext aufgrund eines Rechercheergebnisses an, was natürlich auch zur Nutzungsform Themasure zu rechnen ist. Immerhin noch 8% stießen vermutlich einfach aus Neugier auf elektronische Volltextzeitschriften.

Angebot elektronischer Zeitschriften an der Universitäts- und Landesbibliothek Düsseldorf

(Stand: Dezember 1999)

| <u>Fachgebiet</u> | <u>Anteil</u> |
|-------------------------------------|----------------------|
| Naturwissenschaften / Technik | 48% |
| Medizin / Pharmazie | 18% |
| Sozialwissenschaften | 18% |
| Geisteswissenschaften | 5% |
| Sprach- und Literaturwissenschaften | 3% |
| Sonstiges | 8% |

| <u>Anbieter</u> | <u>Anteil</u> |
|------------------------|----------------------|
| Elsevier | 34% |
| EBSCO | 33% |
| Springer | 11% |
| Kluwer | 10% |
| Wiley | 2% |
| Andere | 10% |

Vergleichende Nutzungsuntersuchung von Print-Zeitschriften und elektronischer Version

Solange Bibliotheken noch nicht großflächig auf das ausschließliche Angebot der elektronischen Version einer Zeitschrift übergegangen sind, wird die Nutzung elektronischer Zeitschriften (mit Ausnahme reiner online-Titel) immer durch das Vorhandensein entsprechender paralleler Print-Abonnements beeinflusst. Die Gründe hierfür liegen auf der Hand: Der Hang zum Traditionellen, Unkenntnis des elektronischen Angebotes, räumliche Nähe zur Zeitschriftenauslage etc. Die Nutzung der Print-Versionen reduziert natürlich die (messbaren) elektronischen Zugriffe. Auf der anderen Seite ist aber auch interessant, ob sich das Nutzerverhalten im konventionellen und im elektronischen Bereich voneinander unterscheiden. Sollte dies nicht generell der Fall sein, wäre das im elektronischen Bereich sehr genau messbare Nutzungsverhalten auf den Print-Bereich übertragbar und würde damit zum ersten Mal verlässliche Aussagen über die Nutzung auch im Print-Bereich ermöglichen.

Als flankierende Massnahme wurde daher über einen Zeitraum von sechs Wochen eine parallele Untersuchung der in der mathematisch-naturwissenschaftlichen Fakultät gehaltenen 152 Elsevier-Zeitschriften durchgeführt. Die Benutzer wurden gebeten, pro Nutzung einen Abreißzettel in eine Sammelbox zu werfen, die täglich geleert und ausgewertet wurde. Zum Vergleich mit den elektronischen Daten wurden die im Untersuchungszeitraum gemessenen Werte mittels im Projekt COMBI¹⁷ erprobter statistischer Verfahren auf ein Jahr hochgerechnet. Bild 8 zeigt das Ergebnis:

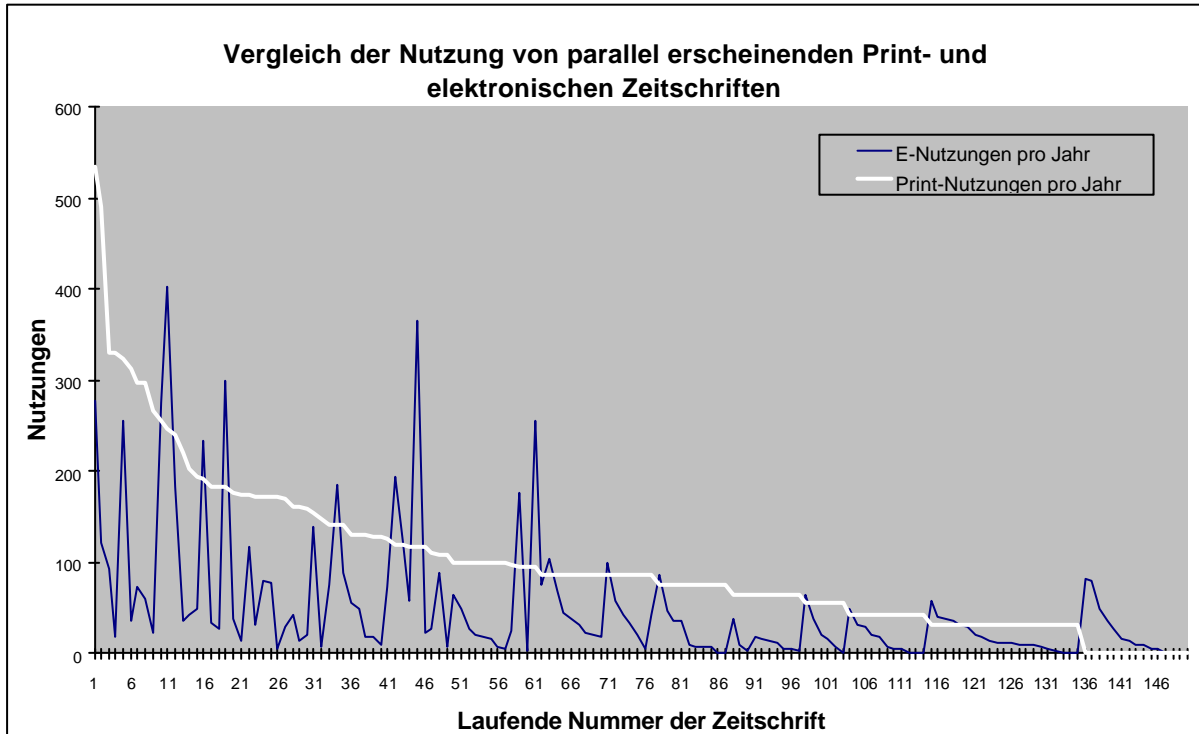


Bild 8: Vergleichende Nutzungswerte zwischen Print- und elektronischer Version

¹⁷ COMBI = Controlling und Marketing für wissenschaftliche Bibliotheken.
<http://www.uni-duesseldorf.de/WWW/ulb/combi.html>

Die helle Kurve zeigt den typischen Verlauf der Nutzungsverteilung eines Zeitschriftenpaketes, beginnend mit dem am meisten genutzten Titeln (Nr. 1). Die dunkle Kurve zeigt die Nutzung des gleichen Titels in elektronischer Form.

Es ist deutlich zu sehen, dass eine generelle Korrelation zwischen beiden Medienformen nicht gegeben ist. Es gibt sowohl viel genutzte Print-Titel, bei denen die elektronische Nutzung hoch ist, aber auch Beispiele, wo sie niedrig ist. Auf der anderen Seite werden im elektronischen Bereich aber auch Spitzen bei Titeln gemessen, die als Print-Version nur wenig gefragt sind.

Nutzung elektronischer Zeitschriften an der Universitäts- und Landesbibliothek Düsseldorf 1999

Die Entwicklung der Akzeptanz elektronischer Zeitschriften soll hier dargestellt werden an der zeitlichen Entwicklung der Volltextzugriffe bzw. der Sessions.

Zu Beginn des Projektes, im Frühjahr 1998, war die Bekanntheit und damit auch die Nutzung des neuen Angebotes noch gering. Durch frühzeitige Werbemaßnahmen konnte bis März 1999 eine erhebliche Steigerung der Nutzung erreicht werden. Zu diesem Zeitpunkt wurde ein Sättigungsniveau erreicht, das sich nur noch durch saisonale Einflüsse nach oben und unten veränderte und daher eine gute Basis für statistische Nutzungsuntersuchungen bildete.

In den ersten drei Monaten 1999 zeigte sich ein deutlicher Anstieg der Sessions, in denen elektronische Zeitschriften aufgerufen worden waren. Ab März 1999 stabilisierte sich die Nutzung zwischen ca. 2600 und 3200 Volltextzugriffen bzw. zwischen ca. 900 und 1100 Sessions. An dieser Entwicklung ist zum einen zu sehen, dass die im Projekt durchgeführten Werbemaßnahmen ab Frühjahr 1999 die gewünschte Wirkung erzielten und zum anderen ist zu vermuten, dass bei ca. 3000 Volltextzugriffen eine Sättigung erreicht war - ein Umstand, der wesentlich zur Authentizität des Untersuchungsergebnisses beiträgt.

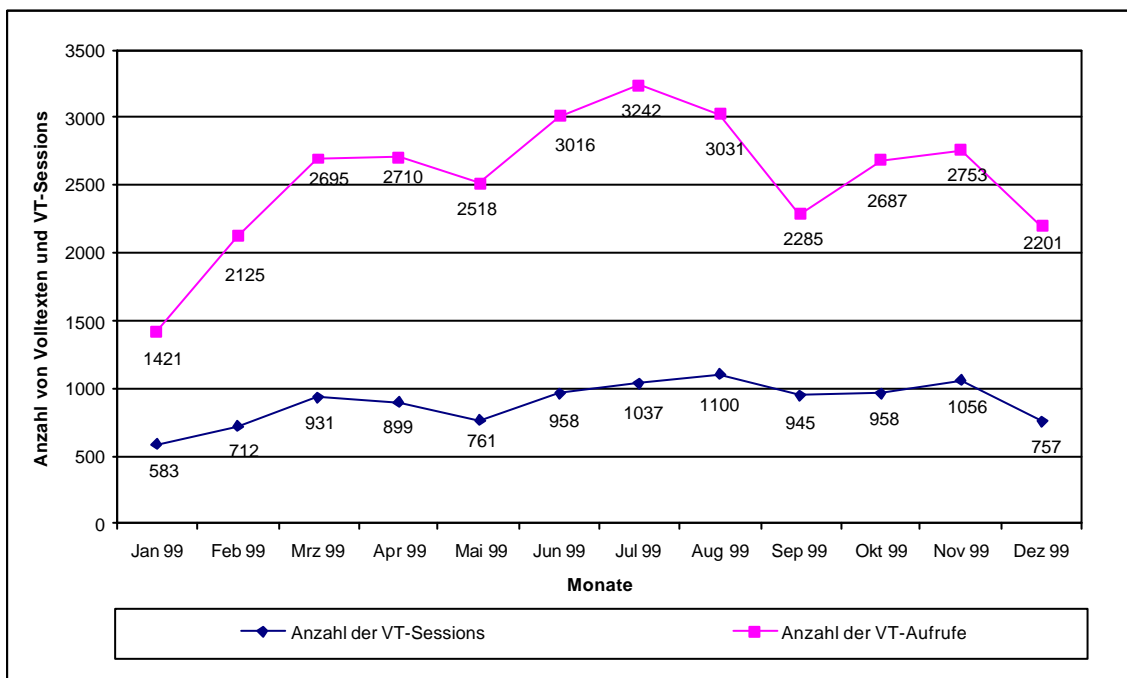


Bild 9: Zeitliche Entwicklung der Nutzung in 1999

Nutzungsbreite des untersuchten Zeitschriftenpaketes an der Universitäts- und Landesbibliothek Düsseldorf 1999

Die erste Grafik gibt einen Überblick darüber, wieviele der insgesamt 1275 untersuchten Titel wie häufig an der ULB Düsseldorf in 1999 genutzt wurden. Es zeigte sich, dass die große Maße der Titel im unteren Nutzungsbereich angesiedelt ist: 240 Titel (19%) wurden überhaupt nicht genutzt, 874 Titel (69%) liegen im Bereich zwischen 1 und 49 Volltextaufrufen pro Jahr, wobei innerhalb dieses Bereiches wiederum die meisten Titel im unteren Nutzungsbereich liegen, wie die zweite Grafik zeigt:

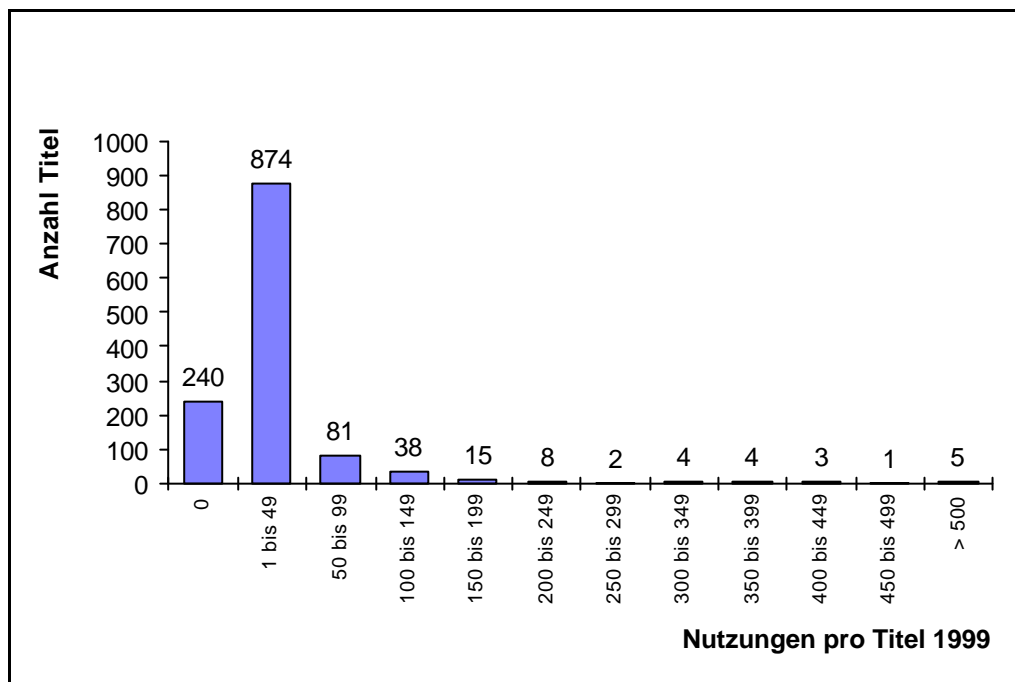


Bild 10: Nutzungsverteilung über das Gesamtangebot von 1275 Zeitschriftentiteln

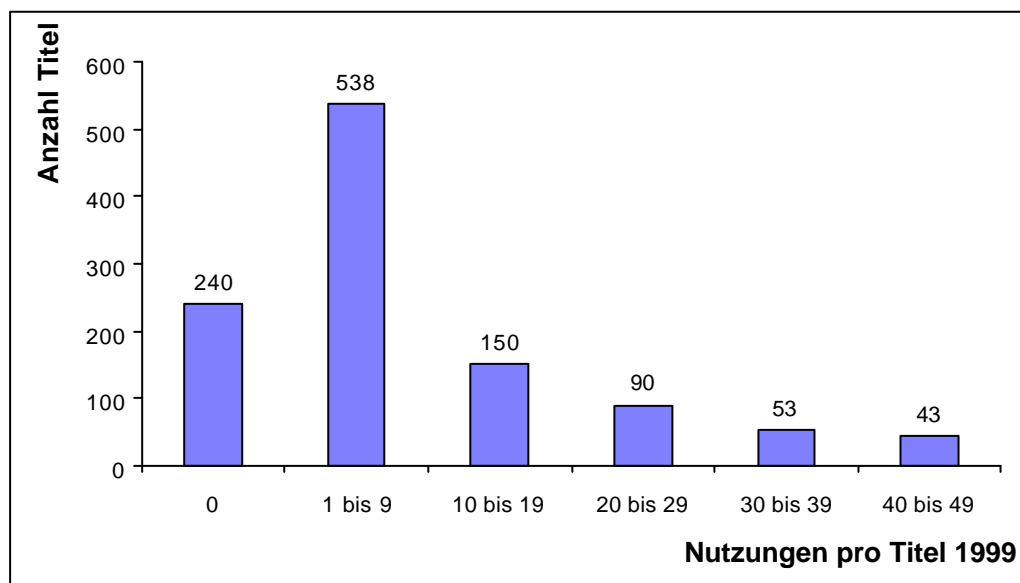


Bild 11: Nutzungsverteilung im Bereich bis 50 Volltextnutzungen pro Jahr

Nutzungsbreite einzelner Titel

Vor dem Hintergrund immer neuer Abbestellaktionen vieler Bibliotheken als Reaktion auf jährlich steigende Zeitschriftenpreise und der damit verbundenen Diskussionen mit den Fakultäten, wurde die Nutzungsbreite der TOP-Titel untersucht. Dabei sollte festgestellt werden, von wievielen verschiedenen Nutzern und in welcher Häufigkeit diese Titel aufgerufen wurden. Stellvertretend für die untersuchten TOP-Ten-Titel wird auf der nächsten Seite das Ergebnis der am meisten genutzten Zeitschrift "Trends in Neurosciences" dargestellt.

Dabei fällt auf, dass nur wenigen der insgesamt 61 Nutzer eine regelmäßige Nutzung dieses Titels zu attestieren ist (Nutzer2 kann nicht gewertet werden, da es sich um den Proxy-Server der Universität Düsseldorf handelt). Im Gegenteil, es fällt auf, dass zwei Drittel die Zeitschrift (wenn auch in vielen Fällen mehrere Artikel) überhaupt nur einmal aufgerufen haben. Gerade in den Fällen mit nur einer Volltextnutzung ist dem sicher in den meisten Fällen ein Retrievalergebnis vorausgegangen, der Aufruf dieses Titels daher eher zufällig.

| | Zugriffe 1999 | Jan | Feb | Mär | Apr | Mai | Jun | Jul | Aug | Sep | Okt | Nov | Dez |
|-----------|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Nutzer 1 | 410 | | | 4 | | 1 | 61 | 1 | 8 | 1 | | 334 | |
| Nutzer 2 | 240 | | | 31 | 14 | 17 | 2 | 32 | 31 | 30 | 37 | 32 | 14 |
| Nutzer 3 | 58 | 2 | 4 | 6 | 2 | 1 | 4 | 6 | 10 | 1 | 4 | 9 | 9 |
| Nutzer 4 | 37 | | 27 | 1 | 1 | | 2 | 5 | | | | 1 | |
| Nutzer 5 | 36 | 1 | 1 | 5 | 13 | 2 | 11 | 2 | | | 1 | | |
| Nutzer 6 | 26 | | 26 | | | | | | | | | | |
| Nutzer 7 | 21 | | 1 | | 9 | | | 2 | 4 | 1 | 2 | 2 | |
| Nutzer 8 | 16 | | | | | | 2 | | 14 | | | | |
| Nutzer 9 | 15 | | | 1 | 2 | 1 | 1 | 7 | | | | 2 | 1 |
| Nutzer 10 | 9 | | | 1 | 2 | 3 | 1 | 2 | | | | | |
| Nutzer 11 | 8 | 8 | | | | | | | | | | | |
| Nutzer 12 | 8 | | 2 | 6 | | | | | | | | | |
| Nutzer 13 | 8 | 2 | 3 | | | | | | 2 | | | 1 | |
| Nutzer 14 | 6 | | | | 2 | | | | 4 | | | | |
| Nutzer 15 | 5 | | 3 | | | | | | 2 | | | | |
| Nutzer 16 | 5 | | 4 | | | | | | 1 | | | | |
| Nutzer 17 | 4 | | | | | | | 3 | | | | | 1 |
| Nutzer 18 | 4 | 1 | | | | | | | 3 | | | | |
| Nutzer 19 | 3 | | | | | | | | | | | | 3 |
| Nutzer 20 | 3 | | | | | | | | | 3 | | | |
| Nutzer 21 | 3 | | | | | | | | | | | 3 | |
| Nutzer 22 | 3 | | | | 2 | | | | | | 1 | | |
| Nutzer 23 | 3 | | | | | | | | | 3 | | | |
| Nutzer 24 | 3 | | | | | | | | | | 2 | 1 | |
| Nutzer 25 | 3 | | | | | | | | | 1 | 2 | | |
| Nutzer 26 | 3 | | | | | | | | | | | | 3 |
| Nutzer 27 | 2 | | 2 | | | | | | | | | | |
| Nutzer 28 | 2 | | | | | | 2 | | | | | | |
| Nutzer 29 | 2 | | | | | | | | 2 | | | | |
| Nutzer 30 | 2 | | | | | | | | | | 2 | | |
| Nutzer 31 | 2 | | | | | | 2 | | | | | | |
| Nutzer 32 | 1 | | 1 | | | | | | | | | | |
| Nutzer 33 | 1 | | 1 | | | | | | | | | | |
| Nutzer 34 | 1 | | | | | | 1 | | | | | | |
| Nutzer 35 | 1 | | | | | | 1 | | | | | | |
| Nutzer 36 | 1 | | | | | | | 1 | | | | | |
| Nutzer 37 | 1 | | | | | | | | | | | 1 | |
| Nutzer 38 | 1 | | 1 | | | | | | | | | | |
| Nutzer 39 | 1 | | | | 1 | | | | | | | | |
| Nutzer 40 | 1 | | | | | | | | | | | | 1 |
| Nutzer 41 | 1 | | | | | | | | 1 | | | | |
| Nutzer 42 | 1 | 1 | | | | | | | | | | | |
| Nutzer 43 | 1 | | | | | | | | | 1 | | | |
| Nutzer 44 | 1 | | | | 1 | | | | | | | | |
| Nutzer 45 | 1 | | | | | | | | | | | 1 | |
| Nutzer 46 | 1 | | 1 | | | | | | | | | | |
| Nutzer 47 | 1 | | | | | | | | 1 | | | | |
| Nutzer 48 | 1 | | | | | | | | 1 | | | | |
| Nutzer 49 | 1 | | | | | | | | 1 | | | | |
| Nutzer 50 | 1 | | | | | | | | | | 1 | | |
| Nutzer 51 | 1 | | | | | | | | 1 | | | | |
| Nutzer 52 | 1 | | | | | | | | 1 | | | | |
| Nutzer 53 | 1 | | | | | 1 | | | | | | | |
| Nutzer 54 | 1 | | 1 | | | | | | | | | | |
| Nutzer 55 | 1 | | | | | | | | | | 1 | | |
| Nutzer 56 | 1 | | | | | | 1 | | | | | | |
| Nutzer 57 | 1 | 1 | | | | | | | | | | | |
| Nutzer 58 | 1 | | | | | | | 1 | | | | | |
| Nutzer 59 | 1 | | | | | | | 1 | | | | | |
| Nutzer 60 | 1 | | | | | | | | | | 1 | | |
| Nutzer 61 | 1 | | | | 1 | | | | | | | | |

Bild 12: Zeitliche Frequentierung des Spitzentitels durch alle aufrufenden Nutzer

Rechnerbezogene Nutzungsverteilung

Neben Erkenntnissen über das Vorgehen bei der Nutzung elektronischer Zeitschriften sind auch Daten über die Anzahl und Struktur der Nutzer von Bedeutung. Um Hinweise darauf zu erhalten, wurde eine rechnerbezogene Auswertung der LogFiles vorgenommen.

Es konnten 1017 Rechner gezählt werden, von denen aus Sessions mit Volltextaufrufen gestartet wurden. Dabei sind insgesamt von 46 Rechnern (5 %) aus sehr ausgedehnte Sessions mit 100 und mehr Volltextzugriffen durchgeführt worden. Von 64 % der Rechner wurden weniger als 10 Volltexte pro Session aufgerufen. Von 18 % der Rechner wurde insgesamt nur ein einziger Volltext aufgerufen. Die folgende Grafik verdeutlicht die Verteilung der Volltextaufrufe auf verschiedene Rechner:

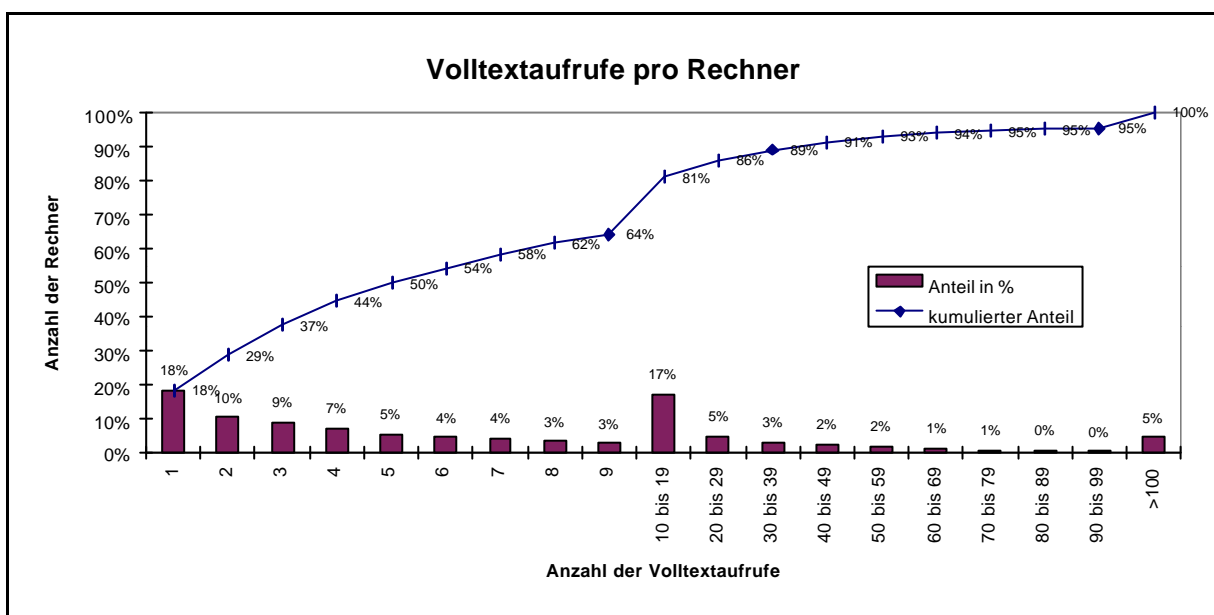


Bild 13: Verteilung der Volltextaufrufe auf Rechner

Diese Auswertung zeigt, dass es eine Reihe von Nutzern gibt, die das elektronische Zeitschriftenangebot kennen und auch nutzen, wobei über die Regelmäßigkeit hier keine Aussage gemacht werden kann. Desweiteren ist anzumerken, dass hinter den hier angenommen Nutzern keine Personen identifizierbar sind, ja noch nicht einmal zu erkennen ist, ob sich hinter einer URL/IP immer dieselbe Person verbirgt, weil viele Aufrufe von PC-Pools innerhalb der Universität oder der Bibliothek kamen.

Nutzung elektronischer Zeitschriften in verschiedenen Wissenschaftsgebieten

Die absoluten Nutzungszahlen waren in den naturwissenschaftlichen Fächern (STM) erwartungsgemäß am höchsten, da in diesem Fachbereich die meisten elektronischen Elsevier-Zeitschriften angeboten wurden. In allen drei Wissenschaftsgebieten lag der Großteil der Zeitschriften bei einer niedrigen Nutzungshäufigkeit (insbesondere unter 10 Nutzungen). Eine hohe Nutzungshäufigkeit (50 Nutzungen und mehr) wiesen dagegen nur wenige Zeitschriften auf, wobei hier deutlich Zeitschriften aus STM-Fächern dominierten.

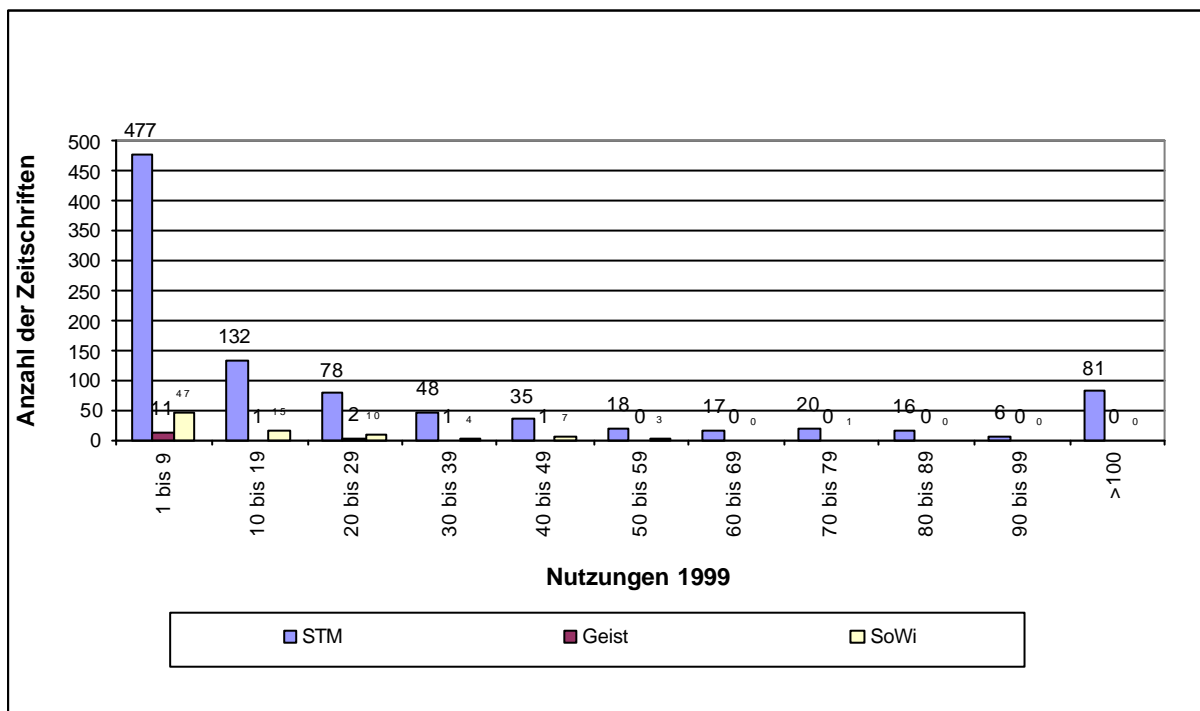


Bild 14: Vergleich der Nutzungshäufigkeiten von Geistes- und Sozialwissenschaften und dem STM-Bereich

Nutzungsanteile verschiedener Jahrgänge

Das der Untersuchung zugrundeliegende Zeitschriftenpaket umfasste die Jahrgänge 1995-1999. Die Aussagefähigkeit wird allerdings dadurch beeinflusst, dass der Jahrgang 1999 über das Jahr 1999 erst sukzessive verfügbar wurde.

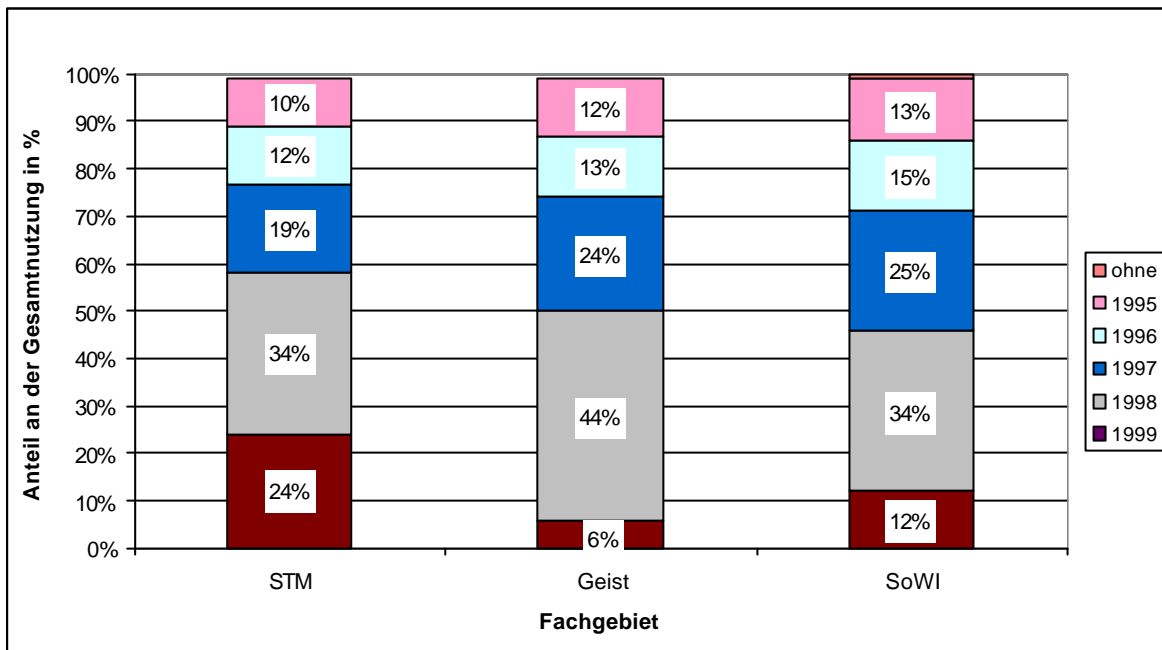


Bild 15: Verteilung der Volltextaufrufe auf Rechner

Bezogen auf verschiedene Jahrgänge zeigte sich bei allen Fächern, dass der Hauptanteil von genutzten Zeitschriften aus 1998 stammt. Bei naturwissenschaftlichen Zeitschriften wurden immerhin auch knapp ein Viertel aus 1999 aufgerufen. Bei den Geisteswissenschaften sowie den Sozial- und Wirtschaftswissenschaften wurden dagegen nur 6 % bzw. 13 % aus dem Jahr 1999 genutzt. Die Aussage, Naturwissenschaftler bevorzugen aktuellere Literatur, bestätigte sich, da 58 % der Nutzungen auf das laufende oder vergangene Jahr gingen, bei den anderen Fachbereichen lag dieser Anteil mit 48 % bzw. 46 % deutlich niedriger. Dagegen konnte die These, geisteswissenschaftliche Fächer benötigen eher ältere Literatur als andere Fächer nicht bestätigt werden. 27 % der Volltextaufrufe geisteswissenschaftlicher Zeitschriften bezogen sich auf die Jahre 1996 und 1995. Bei sozial- und wirtschaftswissenschaftlichen Fächern lag dieser Anteil sogar bei 28 %, bei naturwissenschaftlichen Zeitschriften immerhin bei 22 %. Bezieht man den Jahrgang 1997 noch in die Betrachtung ein, zeigten sich bei den Geisteswissenschaften ein Anteil von 52 % Volltextaufrufe auf ältere Jahrgänge (1995, 1996, 1997) und bei Sozial- bzw. Wirtschaftswissenschaften ein Anteil von 54 %. Die Naturwissenschaften erreichten nur einen Anteil von 41% älterer Literatur. Somit lagen die Geisteswissenschaften bezüglich ihrer Präferenz nach älterer Literatur sogar noch knapp hinter den Sozial- und Wirtschaftswissenschaften.

Nutzung auf Articlebene

Bei der Publikation elektronischer Zeitschriften wurde bisher das Ziel verfolgt, dem Benutzer ein genaues Abbild der Print-Version auf dem Bildschirm zur Verfügung zu stellen. Dies beinhaltet selbstverständlich auch die Kopie der inhaltlichen Struktur, d.h. mehrere Artikel sind zu einem Heft zusammengefasst. Während dieser Heftaufbau bei Print-Zeitschriften eine Voraussetzung für eine effiziente Distribution darstellt, ist er im elektronischen Bereich fast schon obsolet. Lediglich die Zitierfähigkeit wird noch als Argument für die Beibehaltung dieser Struktur angeführt, doch praktisch gesehen werden nicht zuletzt durch den Einsatz mächtiger Retrievalwerkzeuge nur noch einzelne Artikel genutzt - die Einbettung in die Umgebung eines bestimmten Zeitschriftentitels wird zunehmend weniger wichtig.

Vor dem Hintergrund dieser Entwicklung wurden die Nutzungsuntersuchungen im Projekt über die Titlebene hinaus auch auf tiefere Ebenen ausgedehnt. So zeigte z.B. eine Analyse der Hefte, dass durchschnittlich je Heft 2,7 Artikel aufgerufen wurden. Allerdings herrschte dabei eine große Streuung, wie aus folgender Grafik ersichtlich wird:

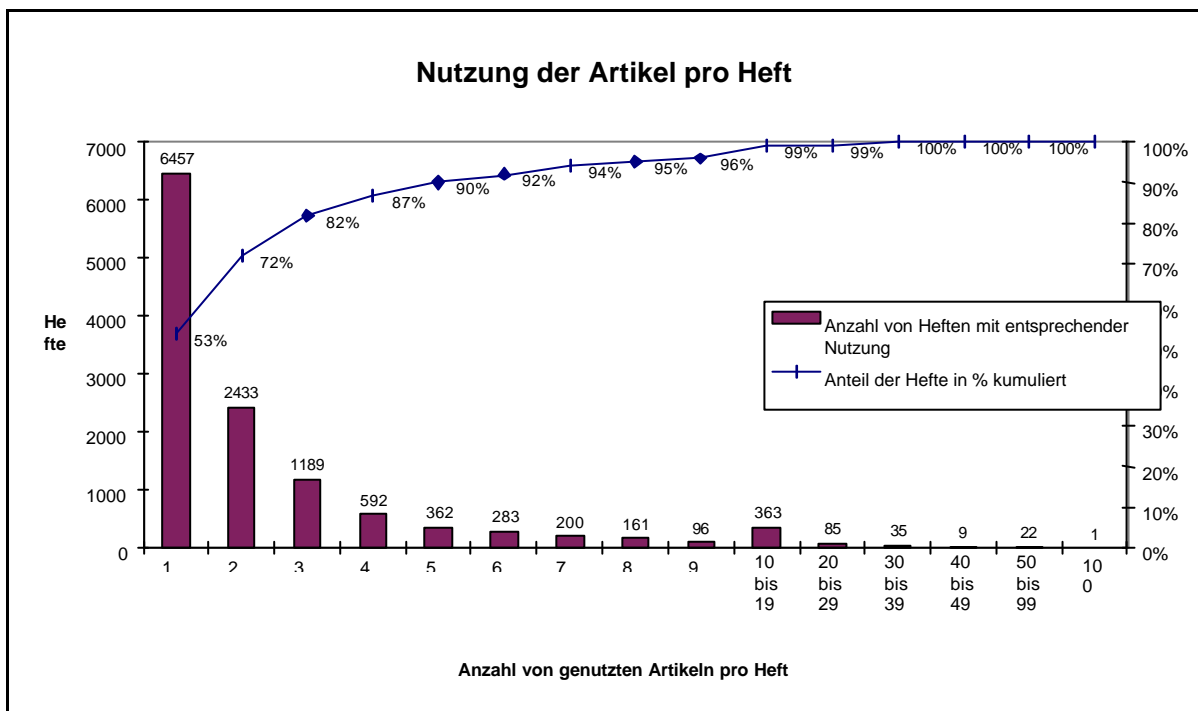


Bild 16: Aus y Hefte wurde x Artikel aufgerufen

Aus 6.457 Hefte (= 53 %) wurde jeweils lediglich einmal ein Artikel genutzt. Dagegen wurden nur aus 504 Hefte (=4%) 10 mal oder häufiger Artikel genutzt. Maximal wurden aus einem einzigen Heft 100 mal Artikel aufgerufen, wobei es sich dabei um 17 verschiedene Artikel handelte, die bis auf einen mehrfach (bis zu 10 Mal) genutzt wurden. Bei Hefte, aus denen 50 mal und häufiger Artikel aufgerufen wurden, sind durchschnittlich 17 Artikel genutzt worden, die durchschnittlich 3,9 mal genutzt wurden.

Auf der Articlebene fallen die Nutzungszahlen noch deutlich schlechter aus. Betrachtet man alle Artikel, die in 1999 auf dem Elsevier-Server vorhanden waren, als großen Pool, wurden nur 2,5 % genutzt (hierbei wurde angenommen, dass den 25.187 Artikelaufrufen ca. 1 Million verfügbare Artikel gegenüberstanden). Mehrfachaufrufe einzelner Artikel sind in dieser Zahl schon enthalten.

Nutzung auf Articlebene

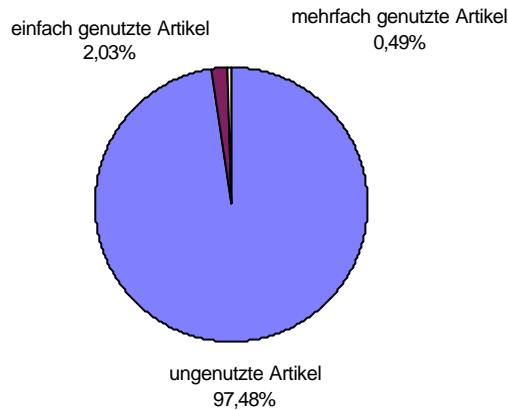


Bild 17: Nutzung der auf dem Server zur Verfügung stehenden Artikel

Im Zusammenhang mit pay-per-view-Modellen wurde schon des öfteren darauf hingewiesen, dass Bibliotheken durch den einmaligen Kauf eines Artikels die Möglichkeit hätten, diesen Artikel durch Zwischenspeicherung mehrfach weiterzugeben und so die Kosten zu senken. Ungeachtet der Tatsache, dass derartige Praktiken in der Regel vertraglich untersagt werden, würden solche Versuche auch kaum einen Sinn machen. Wie die Untersuchung der Mehrfachnutzung von Artikeln belegt, ist es nicht sehr wahrscheinlich, dass ein Artikel mehrfach aufgerufen wird (0,49 % aller verfügbaren Artikel).

Mehrfachnutzung von Artikeln

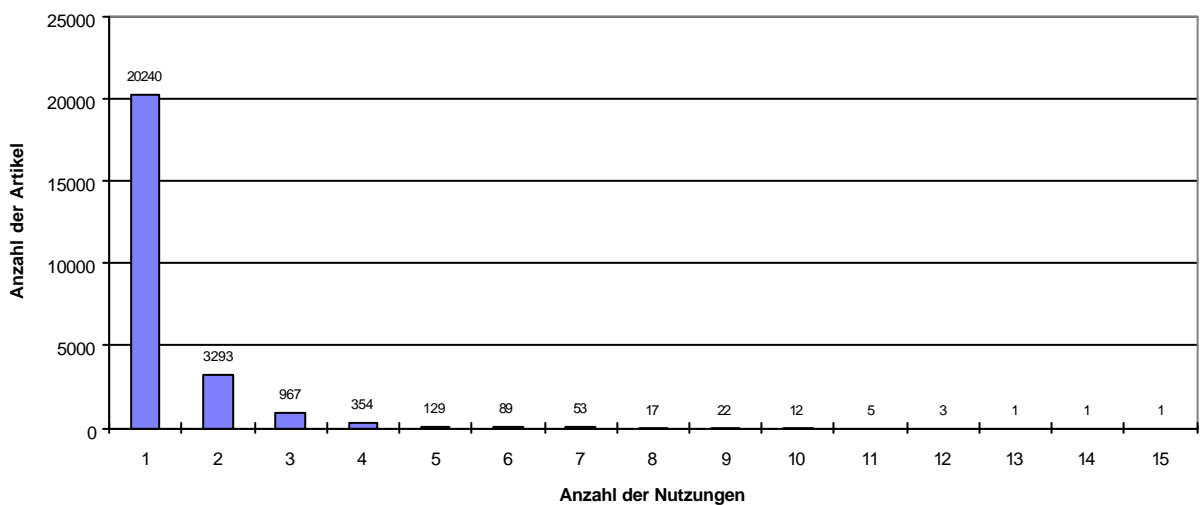


Bild 18: Häufigkeit der Nutzung einzelner Artikel

Von insgesamt 25.187 Artikeln, die während des Jahres 1999 genutzt wurden, sind 81 % nur einmal, 13 % zweimal und 4 % dreimal genutzt worden. Häufiger als dreimal wurden insgesamt nur 687 Artikel genutzt, was einem Anteil von 3 % aller genutzten Artikel entspricht.

Cross- und Additional Access

Distributionsmodelle wie das im Projekt untersuchte Zeitschriftenpaket der Verlagsgruppe Elsevier haben für die einzelne Bibliothek in einem Konsortium den Vorteil, dass, gemessen am Print-Bestand erheblich mehr Zeitschriftentitel angeboten werden können. Zugriffsmöglichkeiten auf Titel, die nicht selbst, aber in wenigstens einer anderen Konsortialbibliothek in der Print-Version gehalten werden, bezeichnet man als Cross Access. Zugriffsmöglichkeiten auf Titel, die überhaupt nicht im Konsortium gehalten werden, als Additional Access. Wie dieser Mehrwert für die Universitäts- und Landesbibliothek Düsseldorf zu bewerten ist, wurde im Projekt untersucht.

1999 wurden an der ULB Düsseldorf 1035 der 1275 elektronischen Elsevier-Zeitschriften mindestens einmal genutzt, wovon 255 Titel in Düsseldorf im Print-Abonnement vorhanden waren, 780 dagegen nicht. Von diesen 780 Titeln wurden 345 Zeitschriften im Konsortium von mindestens einem Mitglied als Print-Version gehalten (Cross Access), 435 dagegen fielen unter Additional Access.

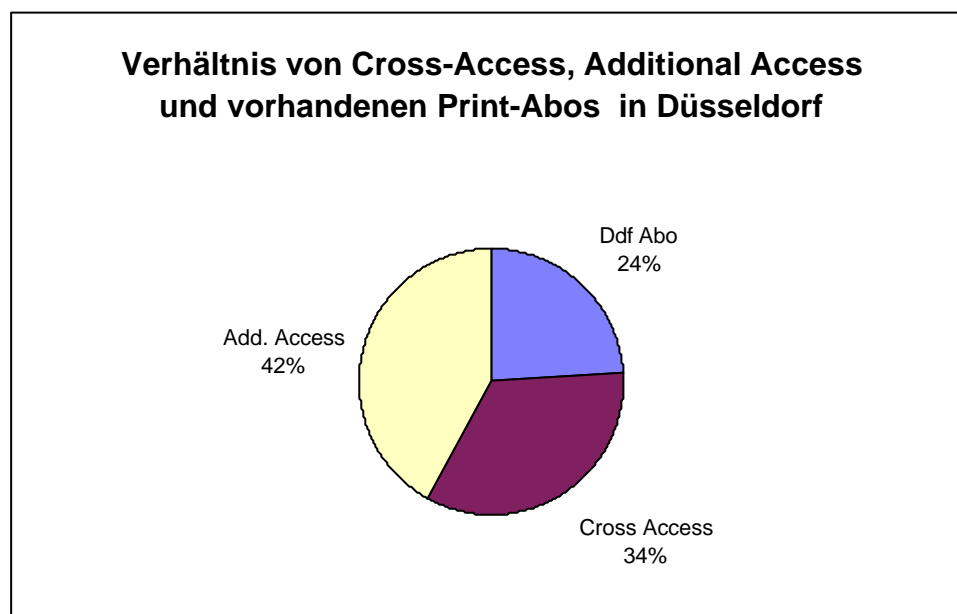


Bild 19: Anteile von Cross- und Additional Access

Durch die Konstruktion, im Konsortium das gesamte Verlagsangebot zu kaufen, scheint der Mehrwert zunächst erheblich zu sein, denn 76% aller genutzten Titel waren in Düsseldorf nicht als Print-Abonnement vorhanden. Betrachtet man allerdings die Nutzungshäufigkeit der Titel in den drei Kategorien, zeigt sich, dass Zeitschriften, die auch im Print-Abonnement gehalten werden, mehr genutzt wurden als die Cross- und Additional-Access Zeitschriften. Die parallel zu den Print-Ausgaben verfügbaren Titel wurden zu 65 % weniger als 50 mal genutzt, zu 35 % häufiger. Im Gegensatz dazu wurden Cross-Access-Titel zu 86 % weniger als 50 mal genutzt und nur zu 14 % häufiger. Bei den Additional-Access-Zeitschriften ist

dieses Verhältnis noch gravierender. Zu 95 % wurden diese Zeitschriften weniger als 50 mal genutzt und nur zu 5 % häufiger.

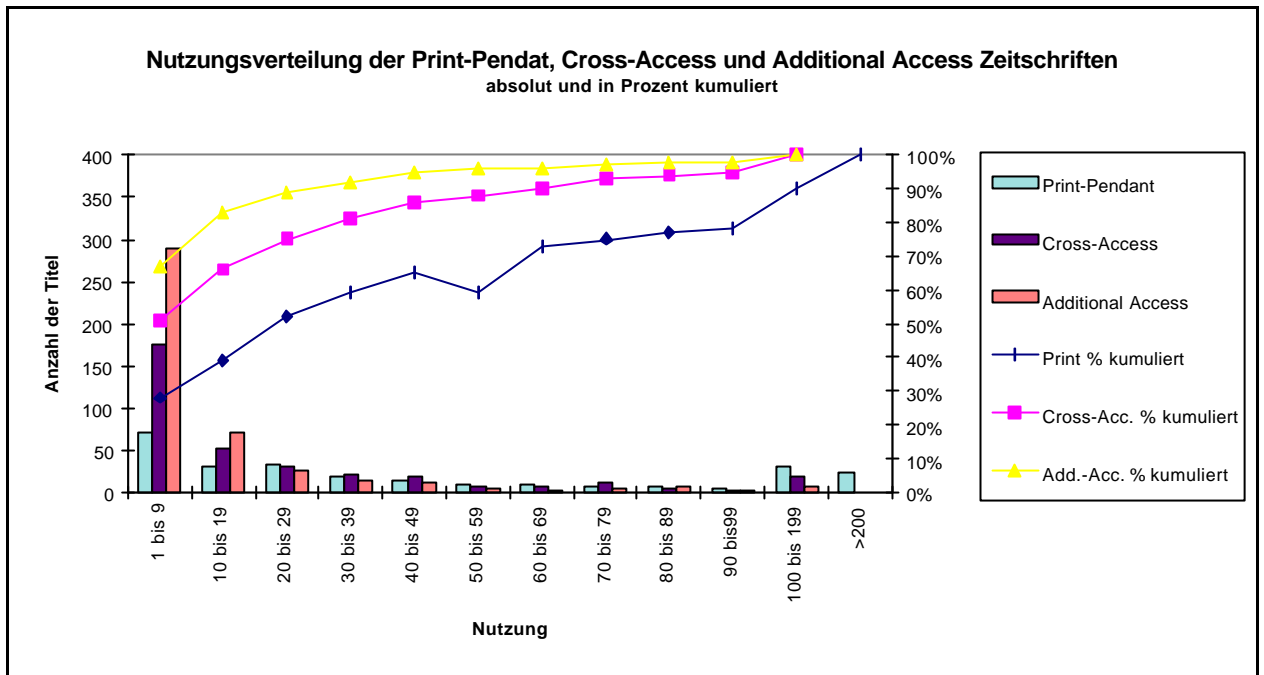


Bild 20: Titelbezogene Nutzungsverteilung bzgl. Cross-, Additional Access und Local Holdings

Auch in absoluten Zahlen machten die zusätzlichen Zugriffsmöglichkeiten durch das Konsortium mit 42,5% aller Volltextaufrufe einen großen Teil aus. Die Häufigkeit der Nutzung einzelner Titel war dagegen gering und die Streuung dafür sehr breit. Auch wenn es allgemein nicht nachzuweisen war, scheint doch eine gewisse Korrelation zu Print-Abonnements gegeben, die dadurch erkennbar wird, dass sich ein fundierter Bestandsaufbau auch im elektronischen Bereich widerspiegelt. Vom Kosten-Nutzen-Verhältnis her betrachtet schneiden Cross- und Additional Access eher schlecht ab und lassen eine pay-per-view-Alternative günstig erscheinen.

Konsortialnutzung

Betrachtet man die Nutzungsverteilung im gesamten nordrhein-westfälischen Konsortium (mit den Teilnehmerbibliotheken UB Bielefeld, ULB Düsseldorf, UB Duisburg, UB Essen, UB Hagen, ULB Münster, UB Siegen und UB Wuppertal), so ergibt sich ein ähnliches Bild, wie für die ULB Düsseldorf allein. Die Zahl der Titel mit Null-Nutzung (34, hier nicht extra ausgewiesen) ist natürlich geringer, da sich die Nutzungsspektren aller Konsortialbibliotheken addieren, die hohe Zahl verhältnismäßig wenig genutzter Titel findet sich aber auch hier. Es bleibt festzuhalten, dass es in einem großen, viele wissenschaftliche Disziplinen abdeckenden Zeitschriftenpaket - wie es hier vorlag - nur verhältnismäßig wenige Titel gibt, denen eine gute Nutzung zu attestieren ist. Das führt dem Schluss, dass nur wenige Titel fächerübergreifend von großem Interesse sind, die große Masse aber nur von wenigen, fachwissenschaftlich Interessierten gelesen werden und eben auch ein großer Anteil, aus welchen Gründen auch immer, kein Interesse bei der Zielgruppe findet.

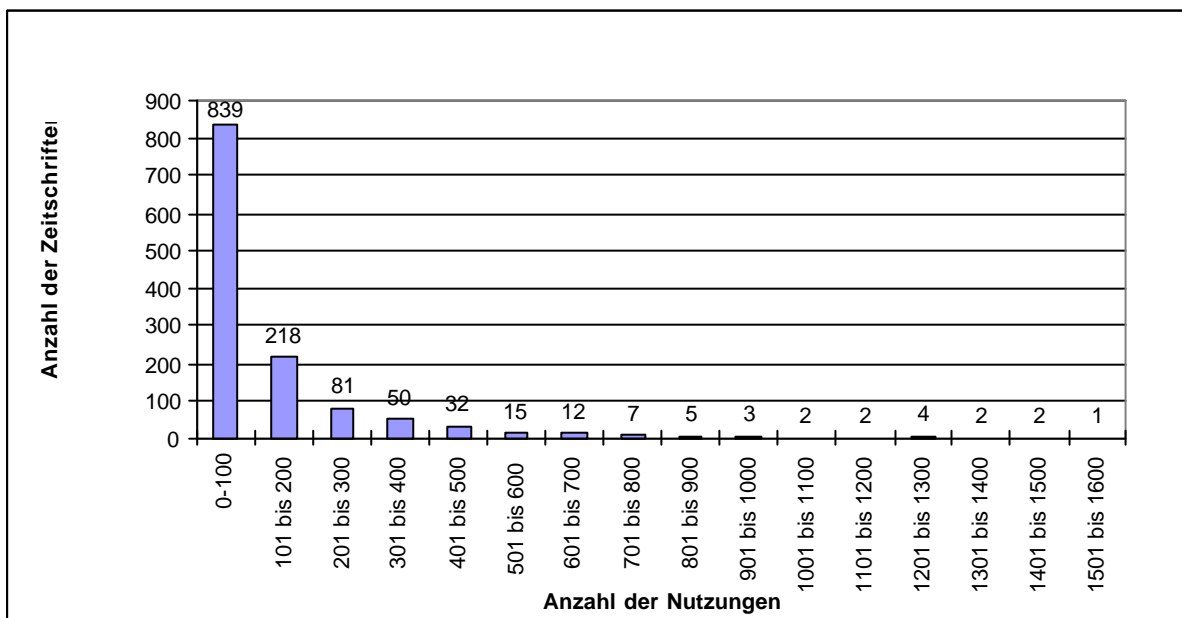


Bild 21: Nutzungsverteilung im Konsortium aller acht Konsortialbibliotheken

Grund- und Spitzenversorgung

Die Differenzierung von Grund- und Spitzenversorgung ist für die DFG ein zentraler Aspekt, schließlich sollte in einem Finanzierungsmodell ausschließlich die Spitzenversorgung gefördert werden.

Das Sondersammelgebietsprogramm kennt die Begriffe Grund- und Spitzenversorgung schon lange; es fällt aber auf, dass es nicht nur keine exakte Definition beider Begriffe gibt, sondern auch, dass sie im Print-Bereich nicht mit Nutzungsintensität eines Titels in Beziehung gesetzt werden. Grundbedarf definiert sich einfach durch die Titel, die über die SSG-Bibliothek hinaus noch an (mehreren) anderen Bibliotheken gehalten werden. Alles andere wäre demnach nur in der SSG-Bibliothek gehaltene Spitzenbedarf, mit dem aber beliebig viele Dokumentlieferanfragen bedient werden können.

Genau das ist aber bei der neuen, elektronischen, Medienform nicht mehr nutzungsunabhängig. Wollte man den Hauptvorteil einer elektronischen Version, nämlich den unmittelbaren Volltextzugriff, erhalten, müsste man eine Nationallizenz abschließen und deren Preis wäre, wenn überhaupt vereinbar, im wesentlichen von der erwarteten Nutzung abhängig.

Auf diese Weise kommt also bei der Untersuchung von Grund- und Spitzenbedarf im elektronischen Bereich der Nutzungsbegriff ins Spiel und es wird deutlich, warum wir uns dieser Frage von der Nutzungsseite her genähert haben.

Zur Grund- und Spitzenversorgung werden im Bereich gedruckter Zeitschriften solche Titel gerechnet, die

- überragend wichtig sind auf verschiedenen Gebieten,
- eine hohe Reputation besitzen,
- eine hohe Benutzungsfrequenz aufweisen,
- an allen Bibliotheken, unabhängig von der lokalen Schwerpunktbildung der Forschung, stark nachgefragt werden.

Wendet man diese Definition auf elektronische Zeitschriften an, bietet sich als Untersuchungsobjekt das gesamte NRW-Konsortium mit seinen acht Universitätsbibliotheken an, die aufgrund ihrer unterschiedlichen Forschungsschwerpunkte schon eine ganz gute repräsentative Basis bilden. Die Forderung nach hoher Benutzungsfrequenz ging in unsere Untersuchung in der Form ein, dass pro Bibliothek nur die Titel berücksichtigt wurden, die mindestens 60 Nutzungen pro Jahr aufwiesen; also durchschnittlich 5 pro Monat.

Mit diesen Parametern kamen wir zu folgendem Ergebnis:

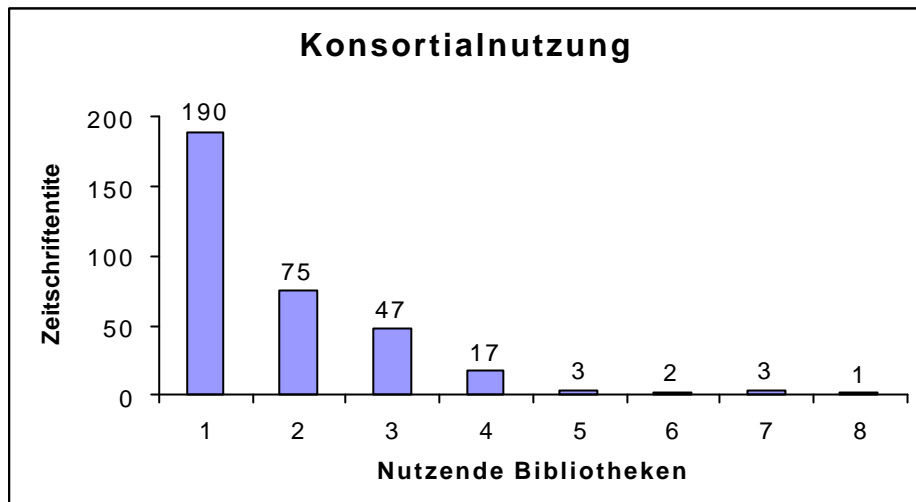


Bild 22: y Titel wurden von x Konsortialbibliotheken genutzt

Insgesamt wurden in den 8 Bibliotheken 338 Titel mindestens 60 mal genutzt. Aber nur einer davon auch 60 mal in allen 8 Bibliotheken. 3 Titel in 7 Bibliotheken usw.

Das heißt ein Grundbedarf würde streng genommen nur aus diesem einen Titel bestehen. Selbst wenn man den Maßstab nicht ganz so eng anlegt und auch noch Titel zum Grundbedarf zählt, die nur in 75% aller Bibliotheken eine hohe Nutzung aufweisen, würden hiernach nur 4 Titel zum Grundbedarf zählen. Geht man ausschließlich von der Nutzungsseite an diese Frage heran, gibt es also keinen überregionalen Grundbedarf, sondern nur einen lokalspezifischen Grundbedarf an den einzelnen Bibliotheken.

Da Spitzenbedarf aufgrund dieser Zahlen nicht zu quantifizieren war, haben wir die umgekehrte Überlegung angestellt und untersucht, wieviele Titel im Konsortium nicht oder nur sehr wenig genutzt werden.

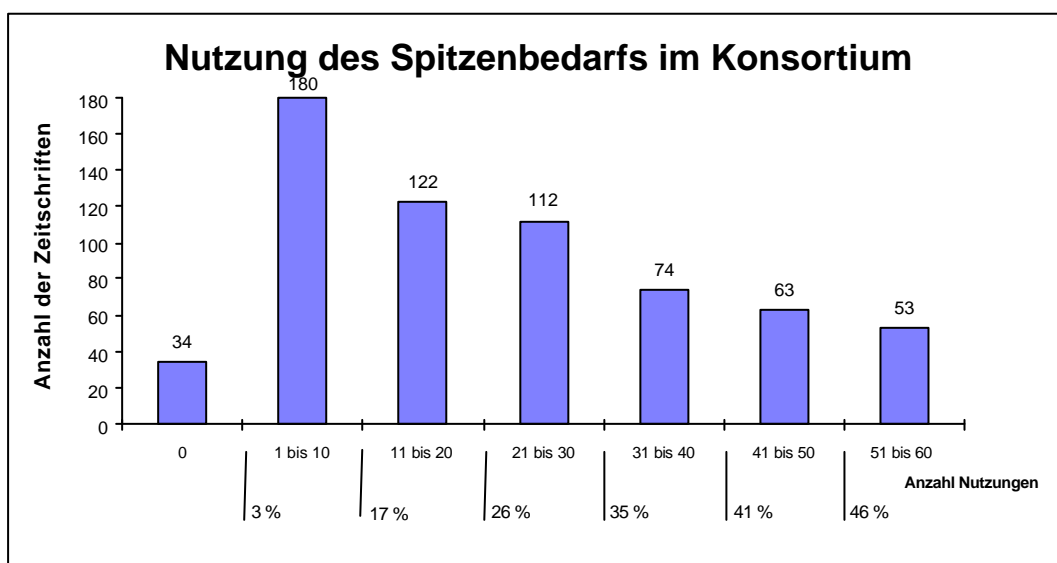


Bild 23: Nutzungsverteilung im Konsortium bei den Titeln, die dem (erweiterten) Spitzenbedarf zugerechnet wurden (bis 60 Nutzungen pro Jahr)

Nach diesen Zahlen stellen die 34 Titel mit Nullnutzung den absoluten Spitzenbedarf dar (wobei das Wort „Bedarf“ hier nicht ganz zutreffend ist). Abhängig davon, bis zu wieviel Nutzungen pro Jahr man noch von Spitzenbedarf sprechen will, erhält man aus diesem Bild aber eine recht gute Mengenangabe.