

Evaluierung des WWW-Angebotes der Bibliothek der Universität Konstanz

Projektarbeit von
Florian Engster

Februar 2002

Titelangaben für dieses Dokument:

Engster, Florian:

Evaluierung des WWW-Angebotes der Bibliothek der Universität Konstanz / von Florian Engster. – Konstanz : Bibliothek der Universität, 2002. – 40 S. : Ill.

Copyright © 2002 by Florian Engster (engster@iuk.hdm-stuttgart.de).

Dieser Text darf zu akademischen Zwecken frei genutzt werden. Eine Nutzung über die Regelungen des Urheberrechts hinaus oder zu kommerziellen Zwecken ist untersagt.

Kurzreferat

Die Nutzung des Web-Angebotes der Bibliothek der Universität Konstanz wurde im Rahmen einer Projektarbeit in der Zeit von September 2001 bis Januar 2002 statistisch erfasst und bewertet. Der Projektbericht stellt die Vorüberlegungen, die Umsetzung und die Ergebnisse dar.

Schlagwörter (SWD)

c.Konstanz / Universitätsbibliothek ; s.Web site ; s.Datenanalyse

Schlagwörter (INFODATA)

- / -

Notation (ACM CCS 98)

H.5.4 ; H.3.5

Inhaltsverzeichnis

1	Zum Verständnis	7
1.1	Das Projekt	7
1.2	Vorgehensweise	7
1.3	Bericht	10
2	Statistik-Legende	11
2.1	Einstieg	11
2.2	Monatsstatistik	12
3	Auswertung „Allgemein“	19
3.1	Generelle Durchschnittswerte	19
3.2	Wochenverlauf	19
3.3	Tagesverlauf	19
3.4	Top-URLs	19
3.5	Ein- und Ausgangsseiten	20
3.6	Nutzerkreise	20
4	Auswertung „Bodensee-Datenbank“	21
4.1	Generelle Durchschnittswerte	21
4.2	Wochenverlauf	21
4.3	Tagesverlauf	21
4.4	Top-URLs	21
4.5	Ein- und Ausgangsseiten	22
4.6	Nutzerkreise	22
5	Auswertung „Datenbankangebot“	23
5.1	Generelle Durchschnittswerte	23
5.2	Wochenverlauf	23
5.3	Tagesverlauf	23
5.4	Top-URLs	23
5.5	Ein- und Ausgangsseiten	24
5.6	Nutzerkreise	24
6	Auswertung „englischsprachiges Angebot“	25
6.1	Generelle Durchschnittswerte	25
6.2	Wochenverlauf	25
6.3	Tagesverlauf	25
6.4	Top-URLs	25
6.5	Ein- und Ausgangsseiten	26
6.6	Nutzerkreise	26
7	Auswertung „Fachinformationsseiten“	27
7.1	Generelle Durchschnittswerte	27
7.2	Wochenverlauf	27
7.3	Tagesverlauf	27
7.4	Top-URLs	28
7.5	Ein- und Ausgangsseiten	28
7.6	Nutzerkreise	28

8	Fazit und Ausblick	29
8.1	Zum Ergebnis	29
8.2	Fortführung	29
9	Anhänge	31
9.1	Software	31
9.2	Anpassung	31
9.3	Datei <code>parse.pl</code>	31
9.4	Datei <code>auswertung.pl</code>	32
9.5	Konfigurationsdatei <code>allg.conf</code>	33
9.6	Konfigurationsdatei <code>bodbibl.conf</code>	33
9.7	Konfigurationsdatei <code>daba.conf</code>	34
9.8	Konfigurationsdatei <code>doklfg.conf</code>	35
9.9	Konfigurationsdatei <code>eng.conf</code>	35
9.10	Konfigurationsdatei <code>epubli.conf</code>	36
9.11	Konfigurationsdatei <code>fi.conf</code>	36
9.12	Konfigurationsdatei <code>intra.conf</code>	37
9.13	Konfigurationsdatei <code>kataloge.conf</code>	38
9.14	Konfigurationsdatei <code>regio.conf</code>	38
9.15	Konfigurationsdatei <code>semcgi.conf</code>	39
9.16	Konfigurationsdatei <code>sis.conf</code>	40

1 Zum Verständnis

1.1 Das Projekt

1.1.1 Allgemeines

Informations- und Dienstleistungs-Angebote im WWW zählen heute auch für Bibliotheken zum Alltag. Da eine professionelle Pflege solcher Angebote je nach Inhalt recht aufwändig ist, haben die Anbieter gerechtfertigtes Interesse daran, dass diese Dienste auch genutzt werden.

Die Besonderheit bei Kommunikation generell und Dienstleistungen speziell, die über Datennetze erfolgen, ist eine relative Anonymität des „Geschäftsvorganges“. Ein zur Verfügung gestelltes Angebot wird zu einem unbestimmten Zeitpunkt von einem unbestimmten Kunden abgerufen.

Das Interesse der UB Konstanz bestand in diesem Fall darin, wissen zu wollen, wie sich die Nutzer des WWW-Angebotes zusammensetzen, welche Teile des Angebotes wie genutzt werden und vor allem, ob sich der Aufwand zur Pflege der arbeitsintensiven Teile vor dem Hintergrund der Ergebnisse überhaupt lohnt.

Ziel des Projektes war daher

1. verschiedene Nutzergruppen zu differenzieren
2. ein Bild der Nutzung des Angebotes zu erstellen
3. dieses Bild vor der Kenntnis der Hintergründe zu interpretieren
4. Rückschlüsse zu ziehen und nach Möglichkeit Verbesserungsvorschläge zu geben

Dabei sollte das in diesem Rahmen entwickelte Konzept auf eine Basis gestellt werden, von der aus es zu einem späteren Zeitpunkt bei Bedarf und Interesse ohne große Umstände übernommen werden kann. Tatsächlich wurde die Automatisierung und dauerhafte Übernahme noch während der Projektarbeit umgesetzt.

1.1.2 Zeitraum

Da die Durchführung im Rahmen eines praktischen Studienseesters erfolgte, war das Fenster prinzipiell auf diesen Zeitrahmen festgelegt. Die Auswertungen erfolgten in wöchentlichen Abständen freitags und beginnen mit den Daten ab dem 01. September 2001. Lücken ergeben sich nur geringfügig durch Urlaub oder sonstige Abwesenheit. Die diesem Bericht zu Grunde gelegten Daten entsprechen der Auswertung vom 25. Januar 2002.

Von diesem Zeitpunkt an erfolgt die Auswertung täglich und automatisiert auf dem Web-Server der Bibliothek und ist über das Intranet tagesaktuell verfügbar:

<http://www.ub.uni-konstanz.de/intranet/sonstiges/>

1.2 Vorgehensweise

1.2.1 Datengrundlage

Datengrundlage für die Erhebung war die Protokoll-Datei des Web-Servers. Hier werden alle Aktionen, die vom Server ausgeführt werden, einheitlich und genau

protokolliert, so dass auch das Verhalten der Nutzer des Web-Angebotes sich hier akribisch genau aufgezeichnet wiederfindet. Die „Kunst“ an der Sache ist die Auswertung dieser Daten nach bestimmten Kriterien.

Zur Auswertung von Server-Logfiles steht eine Reihe von fertigen Programmen zur Verfügung, die diese Aufgabe mehr oder weniger zufriedenstellend erledigen. Eine Gemeinsamkeit dieser Programme ist jedoch der Schwerpunkt auf einer rein statistischen Auswertung der Protokolldatei, die ebenso akribisch erfolgt wie das Protokollieren des Web-Servers. Das Ergebnis ist jedes Mal eine für Menschen lesbare Darstellung der vom Web-Server protokollierten Daten.

Ein für diesen Fall unschöner Gesichtspunkt ist, dass die Protokolldatei auch Daten enthält, die für eine inhaltliche Aussage über die Nutzung des Web-Angebotes keine Rolle spielen bzw. das Ergebnis bei einer Auswertung zwar nicht unbedingt verfälschen, aber unübersichtlicher als notwendig gestalten und zu Lasten der Deutlichkeit der eigentlichen Information gehen.

1.2.2 Vorauswertung

Um den gewünschten Effekt zu erzielen, standen drei Möglichkeiten zur Auswahl:

1. Neuprogrammierung eines geeigneten Analyseprogramms
2. massive Anpassung eines bestehenden Analyseprogramms
3. Datenmanipulation und Nutzung eines bestehenden Analyseprogramms

Die erste und zweite Möglichkeit erschienen auf Grund des zu investierenden Zeitaufwandes nicht gerechtfertigt, zumal bestehende Analyseprogramme auch hier recht zufriedenstellend arbeiten können.

Da somit aber ein unangepasstes Analyseprogramm die Daten als Eingabe erhält, müssen diese zuvor in eine Form gebracht werden, die schließlich zu dem gewünschten Ergebnis führt. Hierzu wird vor dem eigentlichen Analyselauf eine Manipulation an den Daten vorgenommen.

Grafik-Dateien, Stylesheets, JavaScript Zugriffe auf Grafikdateien, Cascading Style Sheets und JavaScript-Dateien werden ignoriert. Diese spielen für die Inhalte des Angebotes eine mindere Rolle, stellen aber einen hohen Anteil am Gesamtvolumen der aufgerufenen Dateien dar, so dass die Zugriffe hierauf im Verhältnis zum Rest ungleich höher wären und sich so verschleiernd auf die Gesamtrelationen auswirken würden

Start-Seite Zugriffe auf die Startseite des WWW-Angebotes, also die Adressen <http://www.ub.uni-konstanz.de/index.htm> bzw. <http://www.ub.uni-konstanz.de/> werden ebenso ignoriert, da vor allem auf den PCs im Benutzerbereich und auch auf den PCs der Mitarbeiter diese Seite als Startseite eingestellt ist. Die Zugriffe hierauf sind zahlenmäßig sehr hoch, stellen aber keine inhaltliche Nutzung des Web-Angebotes dar.

robots.txt Die Datei `robots.txt`, die sich im Wurzelverzeichnis der Web-Dokumente befindet, dient zur Regelung der Zugriffe auf das Web-Angebot durch sogenannte *Robots*, *Crawler* oder *Spider*. Die Verwendung wird

im *Robots Exclusion Standard (RES)* festgelegt¹. Zugriffe hierauf werden ebenfalls nicht beachtet.

Gecachte Inhalte Besuchte Seiten werden von den Browsern in aller Regel in einem lokalen Zwischenspeicher gehalten, dem *Cache*. Um nicht bei jedem erneuten Besuch die Datei komplett laden zu müssen, sendet der Browser oft eine Abfrage, ob sich die Datei auf dem Server gegenüber der Datei im Cache verändert hat. Diese Überprüfung wird mit Hilfe eines „*Head*“-Aufrufes aus dem HTTP-Standard² vorgenommen. Da ja diese Seiten bereits schon einmal ausgeliefert wurden, werden diese Zugriffe ignoriert.

„OK“-Anfragen Letztendlich werden nur noch jene Anfragen weiterverarbeitet, die mit dem HTTP-Statuscode „200“ (= „OK“ für eine erfolgreiche Anfrage) beantwortet wurden. Mögliche andere Codes wären z. B. „404“ für „Seite nicht gefunden“, „304“ für „Seite wurde nicht geändert“, „206“ für „Unvollständiger Inhalt“. Weitere Codes sind aus dem HTTP-Standard ersichtlich.

Hierbei entfallen ebenfalls die parallelen Abrufe ein und der selben Datei durch den gleichen Nutzer: so hat der Browser *Netscape* bei großen Dateien oft den Download begonnen, parallel dazu die gleiche Datei in kleineren Stücken übertragen. Die technischen Zugriffe waren hier größer als die tatsächlichen.

Null-Byte-Antworten Zur Sicherheit werden hier noch sämtliche Einträge verworfen, in denen die Antwort eine Länge von Null Byte hatte, also nichts gesendet wurde.

Suchmaschine Die hauseigene Suchmaschine *ht://dig* erzeugte teils hohe und das Ergebnis verfälschende Zugriffe. Zugriffe dieser Suchmaschine werden in den Auswertungen ab Februar ignoriert.

1.2.3 Datenmanipulation

In einem Aspekt wurden die verfügbaren Daten auch manipuliert, um die Deutlichkeit der hinterlegten Daten zu erhöhen. So wurden die IP-Adressen der Rechner der verschiedenen Benutzergruppen innerhalb der Universität zu jeweils einem fiktiven Rechner zusammengefasst. Als Folge dessen tauchen die Zugriffe von diesen einzelnen Bereichen als Anfragen von *einem* Rechner auf:

benutzer.ub.uni-konstanz.de für Ursprung der Anfragen von den Rechnern in den Buchbereichen der UB,

mitarbeiter.ub.uni-konstanz.de für Ursprung der Anfragen von den Rechnern der Mitarbeiter der UB,

sonstige.uni-konstanz.de für Ursprung der Anfragen aus dem sonstigen Bereich der Universität

Dieses Umstandes muss man sich gewiss sein, wenn es an die Frage der Anzahl der Besucher geht: hinter den nominell einen Rechnern stecken teils mehrere 100 verschiedene einzelne Rechner!

¹vgl. <http://www.robotstxt.org/wc/exclusion.html>

²vgl. RFC 2616: <http://www.ietf.org/rfc/rfc2616.txt>

1.2.4 Protokoll-Analyse

Die Analyse der Protokoll-Datei wurde, neben einer allgemeinen Auswertung, in diverse Teilbereiche des Web-Angebotes aufgeteilt, um die für diese Bereiche jeweils relevanten Daten nicht mit denen anderer Bereiche zu vermischen, wodurch auch hier wiederum ein Plus an Deutlichkeit gewonnen wurde, wie auch einige Aspekte erst hier deutlich werden konnten.

Ausgewertet wurden:

Allgemein Auswertung für das *komplette* Web-Angebot,

Bodensee-Datenbank Auswertung für das Angebot der Euregio-Bodensee-Datenbank,

Datenbanken Auswertung für das Datenbankangebot der UB,

Dokumentlieferung Auswertung für das Angebot der Fernleihe und Subito,

englischsprachige Seiten Auswertung für das Angebot in englischer Sprache,

Elektronische Publikationen Auswertung für das Angebot über KOPS,

Fachinformationen Auswertung für die Angebote der Fachreferenten,

Intranet Auswertung für die Angebote des Intranet,

Kataloge Auswertung für das Katalogangebot (KOALA),

Regionalkatalog Auswertung für das Angebote des Regionalkataloges,

„**SemCGI**“ Auswertung für das Programm zur Abfrage der Semesterapparate,

„**SIS**“ Auswertung für das Angebot des Schlagwortkataloges.

Die Seiten der elektronischen Semesterapparate finden sich ab Februar in der Auswertung der elektronischen Publikationen wieder.

1.3 Bericht

Der vorliegende Bericht soll zunächst die Praxis bei der Aus- und Bewertung des Web-Angebotes darstellen und eine knappe Anleitung zur Pflege der technischen Umsetzung geben. Für die Auswertungen *Allgemein*, *Bodensee-Datenbank*, *Datenbanken*, *englischsprachige Seiten* und *Fachinformationen* finden sich in den weiteren Kapiteln Einzeldarstellungen mit generalisierten Zahlen und zugehörigen Erläuterungen.

2 Statistik-Legende

Da Protokoll-Datei-Analysen nicht immer leicht zu verstehen sind und sich durch die im hier ergebenden Kontext vorab angewandten Datenbereinigungen und -manipulationen noch weitere Kniffe ergeben, soll hier eingangs kurz erläutert werden, wie die jeweiligen Grafiken, Tabellen und Zahlen zu deuten und zu verstehen sind, was sie definitiv aussagen, was sie aussagen können und woraus sich möglicher Weise Fehler ergeben oder ergeben haben.

Dargestellt werden soll dies am Beispiel der Auswertung für das gesamte Angebot.

2.1 Einstieg

Der Einstieg zu jedem Bereich beginnt mit einer Übersicht über die letzten 12 Monate, sofern für jeden Monat Auswertungen vorliegen. Ältere Daten sind in der aktuellen Ansicht nicht mehr verfügbar.

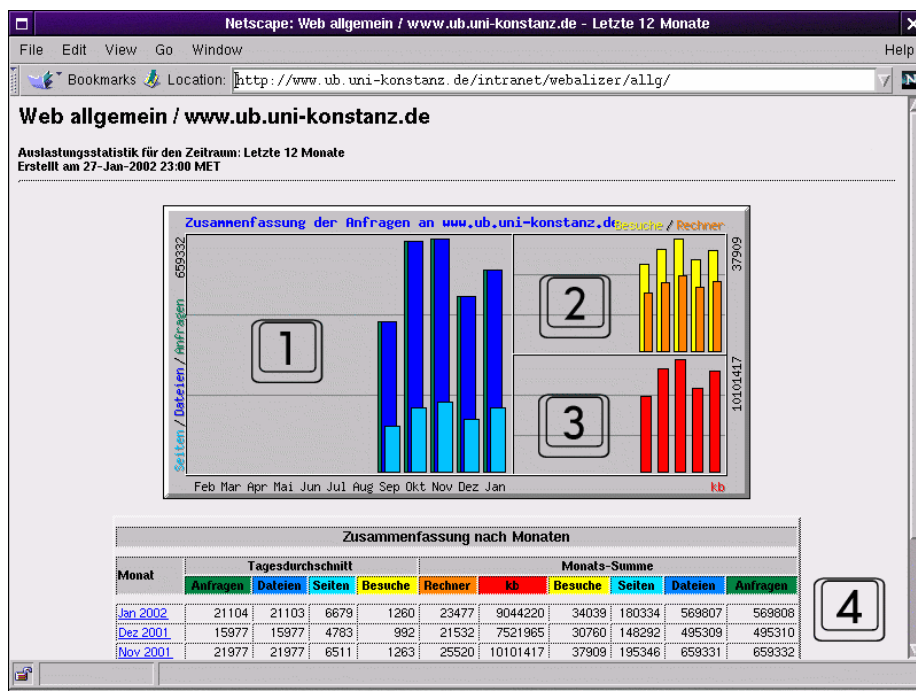


Abbildung 1: Die Einstiegsseite der Auswertung

Die wichtigsten Daten für die einzelnen Monate sind hier ersichtlich und durch die Grafiken visualisiert und entsprechen jeweils den Werten in der darunter liegenden Tabelle. Im Balkendiagramm unter **1** sind die wichtigsten Zugriffsdaten dargestellt:

Anfragen Anfragen oder *Requests* bezeichnen eingehende Anfragen der Browser an eine bestimmte Seite, ohne Berücksichtigung des Ergebnisses. Da wir nur Anfragen auswerten, die mit einem „200“-Statuscode beantwortet

wurden, ist diese Zahl gleich mit der Summe der ausgelieferten Dateien (geringe Abweichungen sind durch Fehler möglich). Ein Unterschied zwischen den beiden Werten würde sich aus der Einbeziehung anderer Statuscodes ergeben, wodurch dieser Wert höher ausfiele.

Dateien Dateien bezeichnet die ausgelieferten Inhalte, also das, was vom Web-Server tatsächlich an den Browser geschickt wird und stellt somit die eigentliche Nutzung dar.

Seiten Mit Seiten sind Dateitypen definiert, die für diese Auswertung gesondert behandelt werden sollen: aktuell umfasst diese Definition Dateien mit den Endungen htm, html und pdf. Der Inhalt dieser Dateien ist *statisch*. Den Unterschied zwischen den Werten der Dateien und Seiten machen die CGI-Skripte, also die *dynamischen Inhalte* aus: in diesem Fall die Anfragen an den Katalog.

Im Balkendiagramm unter [\[2\]](#) sind die Beziehungen zwischen einzelnen Besuchen und Rechnern dargestellt:

Besuche Ein Besuch oder auch „Sitzung“ umfasst ein aneinanderhängendes Abrufen mehrerer Seiten des Angebotes von einem Rechner aus mit dazwischen liegenden Pausen, die nicht länger als 30 Minuten dauern. Eine längere Unterbrechung wird als erneuter Besuch gewertet. Da vor allem an den Rechnern im Buchbereich der UB ständig die Nutzer wechseln, lässt sich daher nicht sagen, wie viele *Personen* auch dahinter stecken. Auch liegen die Unterbrechungen hier in aller Regel unter einer halben Stunde, so dass auf einen Besuch auch gut mehrere Benutzer fallen können.

Rechner Hinter Rechnern steht die Zahl an normaler Weise einzelnen Rechnern, von denen aus eine Abfrage getätigt wurde. Da die Rechner der Universität Konstanz jedoch als zusammen genommen drei Einzelrechner eingerechnet werden, entspricht diese Zahl nicht der tatsächlichen.

Das Balkendiagramm unter [\[3\]](#) stellt dem gegenüber noch die Anzahl der übertragenen Kilobytes, also der Datenmenge dar. In der Regel ist der Kurvenverlauf der Anfragen und der Datenmenge deckungsgleich, Abweichungen können jedoch interessante Hinweise geben: z. B. das Abrufen großer Datenmengen bei kleinen Zugriffszahlen.

Sämtliche Daten sind als Zahlen nochmals in der Tabelle unter [\[4\]](#) hinterlegt, nach Tagesdurchschnitten und Monatssummen aufgeteilt, die Gesamtsummen jeweils unten aufsummiert (nicht im Bild).

2.2 Monatsstatistik

Die Monatsstatistik enthält die Daten in feinerer Auswertung nach:

- Tages-Statistik
- Stunden-Statistik
- Zugriffszahlen
 - nach Anzahl der Zugriffe

- nach übertragener Datenmenge
- Eingangsseiten
- Ausgangsseiten
- besuchende Rechner
 - nach Anzahl der Zugriffe
 - nach übertragener Datenmenge
- über HTTP authentifizierte Benutzer
- Herkunftsländer der Anfragen

2.2.1 Übersicht

Die erste Tabelle der Seite stellt nochmals eine Übersicht zu den Gesamtdaten des Monats dar und listet für die ausgewerteten Daten die Summen, Durchschnitts- und Maximalsumme, sowie die Statuscodes auf. Die Angabe letzterer sollte sich auf „Code 200 – OK“ beschränken. Sollte in den Protokolldateien selbst jedoch ein Fehler stecken, wie hier im Beispiel, so tauchen diese Werte ebenso auf.

The screenshot shows a Netscape browser window with the following content:

Location: http://www.ub.uni-konstanz.de/intranet/webalizer/allg/usage_200201.html

[Tages-Statistik](#) | [Stunden-Statistik](#) | [URLs](#) | [Eingang](#) | [Ausgang](#) | [Rechner \(IP-Adressen\)](#) | [Benutzer](#) | [Länder](#)

Monats-Statistik für Januar 2002		
Summe Anfragen	569808	
Summe Dateien	569807	
Summe Seiten	180334	
Summe Besuche	34039	
Summe kb	9044220	
Summe unterschiedlicher Rechner (IP-Adressen)	23477	
Summe unterschiedlicher URLs	3862	
Summe unterschiedlicher Benutzer	5	
	Schnitt Maximum	
Anfragen pro Stunde	879	5394
Anfragen pro Tag	21104	32084
Dateien pro Tag	21103	32083
Seiten pro Tag	6679	10419
Besuche pro Tag	1260	1766
Kb pro Tag	334971	696466
Anfragen nach Status-Code		
Unbekannter Antwort-Code	1	
Code 200 - OK	569807	

[Tages-Statistik in Monat Januar 2002](#)

Abbildung 2: Monatsstatistik Übersicht

2.2.2 Tagesstatistik

Die Grafik für die Tagesstatistik enthält in ihrer Gestaltung die gleichen Informationen wie die Einstiegsseite, jedoch tagesweise in Spalten angeordnet. Die tabellarische Auflistung enthält zudem Prozentwerte für die einzelnen Daten, die sich auf die Gesamtsummen für den Monat (so weit ausgewertet) beziehen.

2.2.3 Stundenstatistik

Die Stundenstatistik gibt den Verlauf über einen einzelnen Tag hinweg auf die Stunden verteilt wieder, jedoch nur für die Aspekte Anfragen/Dateien und Seiten. Die tabellarische Auflistung trennt zwischen der Schnittsumme, die innerhalb der angegebenen Stunde gültig ist, und der Summe, die insgesamt für diese Stunde im gesamten Monat aufgetreten ist.

2.2.4 Zugriffszahlen

Die Liste der Zugriffszahlen stellt quasi den „*High-Score*“ der Adressen dar, die vom Server ausgeliefert werden. Die Auswertung erfolgt in zwei Tabellen: in der ersten sind die Adressen nach der Summe der Zugriffe geordnet, in der zweiten nach der Summe der Datenmenge (Kilobytes). Grau unterlegte Felder stellen Zusammenfassungen dar: sie besitzen zwar eine eigene Zeilennummer, der angegebene Wert ist allerdings eine Summe aus den restlichen Werten. Im Beispiel sind sämtliche CGI-Skripte aufgerechnet und stehen insgesamt an erster Stelle, die einzelnen Skripte jedoch verteilen sich auf die gesamte Tabelle. Teilweise schleichen sich deutliche Fehler bei den Zusammenfassungen ein, weswegen man diese Zahlen immer mit Vorsicht genießen muss.

Unter der Überschrift „*Top 100 von 3862 URLs*“ ist zu verstehen, dass in der Tabelle 100 einzelne URLs (Adressen) angegeben werden, insgesamt aber 3862 verschiedene URLs in den Protokolldateien aufgetreten sind. Die Überschriften der anderen Tabellen lesen sich entsprechend.

2.2.5 Eingangs-/Ausgangsseiten

Maßgeblich für die Staffelung der Ein-/Ausgangsseiten ist der Wert der Besuche. Nochmals zur Verdeutlichung: als Besuch gelten die Abrufe *eines* Rechners, die zeitlich nicht länger als 30 Minuten auseinander liegen dürfen. Da in der Vorwertung der Daten die Zugriffe auf die Startseite verworfen werden, sind hier die Seiten aufgelistet, die ein Besucher *nach* seinem Einstieg auf das Web-Angebot besucht hat. Wenn es sich bei der Auswertung um einen Teilbereich des Web-Angebotes handelt, so trifft diese Zahl auf die erste besuchte Seite in diesem Teilbereich zu. In Abhängigkeit davon, wie diese Seite vom Rest des Angebotes aus verlinkt ist, ist auch eine Aussage darüber möglich, ob ein Benutzer erst während eines Besuches darauf gelangt ist, oder direkt von einem externen Angebot kam.

Die Aussage der Ausgangsseiten ist analog: hier werden diejenigen Seiten aufgelistet, die ein Besucher zuletzt in diesem Teilbereich abgerufen hat.

Top 100 von 3862 URLs					
#	Anfragen	%	kb	%	URL
1	361547	65.45%	2914726	32.23%	/cgi-bin/
2	211406	37.10%	1821606	20.14%	/cgi-bin/ac/www25/allearo.pl
3	91294	16.02%	653961	7.23%	/cgi-bin/ac/www25/regsrch.pl
4	38591	6.77%	125539	1.39%	/cgi-bin/kontoinf.pl
5	19657	3.45%	181164	2.00%	/koala/
6	15192	2.67%	182864	2.02%	/koala/recherche.htm
7	9625	1.69%	297410	3.29%	/cgi-bin/ac/www25/acindex2.pl
8	6971	1.22%	15467	0.17%	/cgi-bin/ac/www25/asitcgi.pl
9	6262	1.10%	24097	0.27%	/cgi/v3-mysql/kops/ergebnis.html
10	5808	1.02%	49115	0.54%	/cgi-bin/ac/www25/siq.pl
11	5415	0.95%	102424	1.13%	/vfb.htm
12	4934	0.87%	18284	0.20%	/cgi-bin/fwbtest
13	4884	0.86%	76828	0.85%	/dt-suchm.htm
14	4826	0.85%	84974	0.94%	/Kataloge.htm
15	4402	0.77%	46076	0.51%	/koala/rechzdb.htm
16	3737	0.66%	23732	0.26%	/cgi-bin/ac/www25/allearo_regio.pl
17	3242	0.57%	14123	0.16%	/vti_bin/shhtml.exe/search.htm
18	2852	0.50%	22126	0.24%	/daba/dbsa.htm
19	2691	0.47%	18667	0.21%	/links.htm
20	2520	0.44%	16639	0.18%	/epubli.htm
21	2513	0.44%	21685	0.24%	/cgi-bin/v3-mysql/v13/ergebnis_kops.html
22	2425	0.43%	46808	0.52%	/koala/diff.htm
23	2248	0.39%	8057	0.09%	/search.htm
24	2069	0.36%	17855	0.20%	/sis/
25	2013	0.35%	17743	0.20%	/Dokulief.htm

Abbildung 3: Liste der Zugriffszahlen nach Anfragen

2.2.6 Besuchende Rechner

Die Liste der besuchenden Rechner gibt wieder, wie viele Anfragen, bzw. wie viele Kilobytes von den jeweiligen Rechnern abgerufen wurden. Die grau unterlegten Einträge sind dabei wieder Zusammenfassungen, die hier jedoch die Domains bis zur zweiten Stufe zusammenfassen und so eine recht gute Aussage darüber treffen, wie sich die Nutzer verteilen. Auch sind Suchmaschinen hier gut erkennbar.

Die Namen der Rechner werden erst nachträglich bei der Auswertung über ihre IP-Adressen aufgelöst. Bei diesem Verfahren kann es zu Timeouts kommen, so dass einige Adressen, die eventuell auflösbar wären, nicht aufgelöst sind.

Als externe Benutzer treten allerdings auch Studierende und sonstige Universitätsangehörige auf, die sich via Modem über den Zugang eines Internet-Providers einwählen!

2.2.7 Über HTTP authentifizierte Benutzer

Die Tabelle gibt die Liste der Benutzer wieder, die sich über HTTP authentifiziert haben: sie ist für die Auswertung an sich nicht von Interesse.

2.2.8 Herkunftsländer der Anfragen

Die Bezeichnung „Länder“ ist hier abstrakt zu verstehen: ausgewertet werden für diese Darstellung sowohl die Top-Level-Domains der Länder, als auch die generischen TLDs wie .com, .org, etc.

Netscape: Web allgemein / www.ub.uni-konstanz.de - Januar 2002

Location: http://www.ub.uni-konstanz.de/intranet/webalizer/allg/usage_200201.html

Top 25 von 978 Eingangsseiten Gesamt

#	Anfragen	Besuche	URL
1	5415	5828	/vib.htm
2	19657	4379	/koala/
3	4884	3844	/kt-suchm.htm
4	4826	1737	/Kataloge.htm
5	909	894	/fiant/Internet.htm
6	877	812	/i/pha/tittle.html
7	15192	626	/koala/recherche.htm
8	6262	569	/cgi/v3-msgl/kops/ergebnis.html
9	919	527	/cgi-bin/v3-msgl/kops/ergebnis.html
10	563	525	/i/ura/s/bd/
11	2691	505	/Links.htm
12	671	470	/reagio/
13	591	461	/bocdb/
14	2520	432	/epubl.htm
15	600	426	/i/spr/
16	608	381	/i/pha/ueber.html
17	2852	355	/daba/daba.htm
18	613	351	/oeffnung.htm
19	2248	334	/search.htm
20	365	315	/i/arc/qsalt.htm
21	2013	310	/Dokulief.htm
22	403	294	/i/ger/
23	404	289	/i/spr/links.htm
24	959	280	/cgi-bin/v3-msgl/v13/frontdoor.html
25	398	259	/i/pha/onlinez.html

Abbildung 4: Liste der Eingangsseiten

Netscape: Web allgemein / www.ub.uni-konstanz.de - Januar 2002

Location: http://www.ub.uni-konstanz.de/intranet/webalizer/allg/usage_200201.html

Top 100 von 23477 Rechnern (IP-Adressen)

#	Anfragen	Dateien	kb	Besuche	Rechnername
1	388276	388275	4594291	120	uni-konstanz.de
2	239093	239092	2299720	95	benutzer.ub.uni-konstanz.de
3	99788	99788	994951	158	sonstige.uni-konstanz.de
4	49395	49395	1299620	119	mitarbeiter.ub.uni-konstanz.de
5	13273	13273	225469	301	mcbone.net
6	7641	7641	99213	240	t-dialin.net
7	6862	6862	34645	282	uni-karlsruhe.de
8	6509	6509	130603	5	[211.119.133.189
9	6484	6484	25465	284	ubkaa4.ubka.uni-karlsruhe.de
10	3919	3919	87055	306	t-online.de
11	3423	3423	34143	176	fh-konstanz.de
12	3194	3194	297319	4	[211.119.133.162
13	2765	2765	72484	301	aol.com
14	2273	2273	20470	128	sgt2-16-2.mcbone.net
15	2104	2104	25762	76	ba-ravensburg.de
16	1932	1932	17235	110	sgt2-15-2.mcbone.net
17	1903	1903	19966	107	mediaways.net
18	1745	1745	15505	107	sgt2-16-1.mcbone.net
19	1658	1658	17399	116	sgt2-15-1.mcbone.net
20	1642	1642	22045	55	risc02.wi.ba-ravensburg.de
21	1527	1527	36791	336	inktom.com
22	1056	1056	25314	27	googlebot.com
23	1030	1030	33805	12	[66.77.74.187
24	954	954	28659	13	[66.77.74.183

Abbildung 5: Liste der besuchenden Rechner

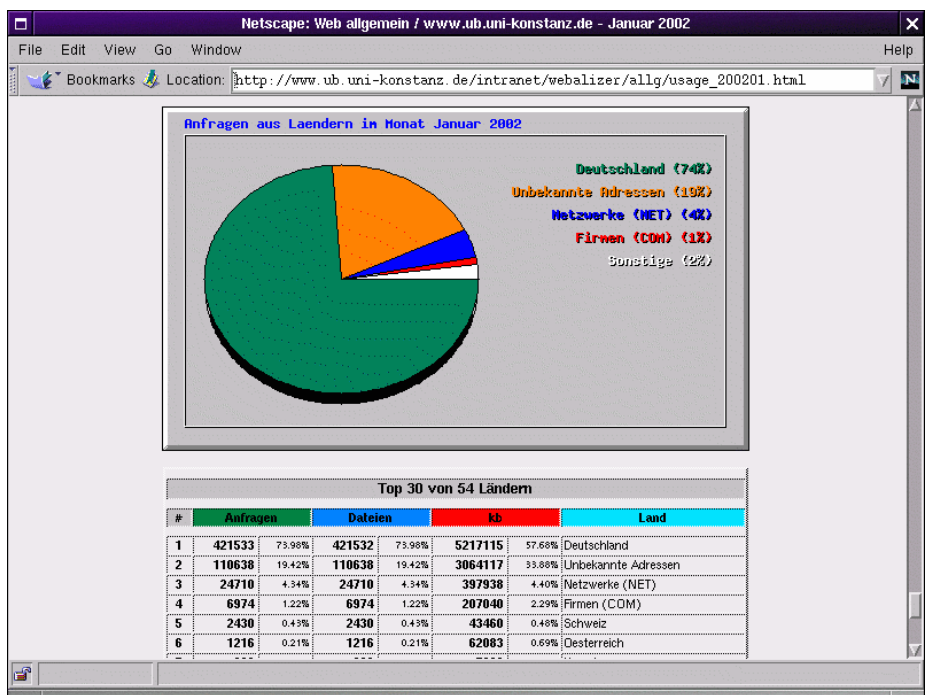


Abbildung 6: Herkunftsländer der Anfragen

3 Auswertung „Allgemein“

Die Auswertung „Allgemein“ umfasst das komplette Web-Angebot der Bibliothek und gibt somit eine Übersicht über die Akzeptanz und Nutzung aller Inhalte, vor allem auch im Bezug zueinander.

3.1 Generelle Durchschnittswerte

Über die Monate betrachtet werden im Durchschnitt pro Besuch 17,43 Dateien abgerufen. Diese überaus hohe Zahl kommt vor allem durch die Katalogabfragen zu Stande. Der sonstige Mittelwert liegt grob gemessen bei ca. 3 Dateien pro Besuch.

Monatsdurchschnitt		
<i>Abrufe ca.</i>	<i>Besuche ca.</i>	<i>Verhältnis</i>
584 000	33 500	17,43 : 1

Tagesdurchschnitt	
<i>Abrufe ca.</i>	<i>Besuche ca.</i>
19 000	1 100

3.2 Wochenverlauf

Die hier erhobene Nutzung der Seiten relativ zu Werktagen und Wochenende gilt als Vergleich für die restlichen Auswertungen. Die Verteilung von 80 % auf die Werktage und 20 % auf das Wochenende stellt sich so fast immer wieder dar.

<i>Wochentage</i>	<i>Anteil ca.</i>
Mo. – Fr.	80 %
Sa. – So.	20 %

3.3 Tagesverlauf

Um die Nutzung über den Tag hinweg zu erfassen wurden drei Zeitfenster betrachtet: die Zeit vor 08:00 Uhr, also vor Beginn der Vorlesungen und vor Beginn der Arbeitszeit, die Zeit zwischen 08:00 und 19:00 Uhr, die mit der Kernzeit von Vorlesungen und Arbeitszeit zusammen fällt, und die Zeit nach 19:00 Uhr.

Die Spitzen der Nutzung liegen in der Regel in der Zeit zwischen 08:00 und 19:00 Uhr.

<i>Zeitfenster</i>	<i>Anteil ca.</i>
00:00 – 08:00 Uhr	3 %
08:00 – 19:00 Uhr	86 %
19:00 – 00:00 Uhr	11 %

3.4 Top-URLs

In der Liste der anteilmäßig am häufigsten abgerufenen Angebote führen die Angebote des KOALA-Systems (Kataloge, Servicefunktionen). Mit 5 % Anteil

stehen auf zweiter Stelle die Fachinformationsseiten und das Angebot an Datenbanken immerhin an dritter Stelle.

<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	40 %	KOALA
2	5 %	Fachinformationen
3	4 %	Datenbanken
4	3 %	Suchmaschinen, VLB, Links, ...
...		

3.5 Ein- und Ausgangsseiten

Die Eingangsseiten der Besuche stellen hier die Seiten *nach* dem Besuch der Index-Seite dar, bzw. die Seiten, auf die über Bookmarks oder externe Links eingestiegen wird.

Eingangsseiten		
<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	13 %	VLB
2	11 %	Suchmaschinen
3	10 %	KOALA
4	5 %	Kataloge
...		

Ausgangsseiten		
<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	12 %	VLB
2	10 %	Suchmaschinen
3	8 %	KOALA
4	2 %	Kataloge
...		

3.6 Nutzerkreise

Bei Verteilung der Nutzung auf verschiedene Nutzerkreise, Universitäts- und externe Nutzer, entfallen grob drei Viertel auf Zugriffe von innerhalb der Universität, grob die Hälfte davon stammt von den Rechnern im Buchbereich.

<i>Nutzerkreis</i>	<i>Anteil ca.</i>
uni-konstanz.de	71 %
<i>davon UB-Mitarbeiter</i>	8 %
<i>davon UB-Nutzer</i>	46 %
<i>davon sonstige Uni</i>	17 %
extern	29 %

4 Auswertung „Bodensee-Datenbank“

Die Auswertung „Bodensee-Datenbank“ umfasst das Angebot der Seiten unter /boddb/ sowie die Katalog-Skripte.

4.1 Generelle Durchschnittswerte

Über die Monate betrachtet werden im Durchschnitt pro Besuch 2,5 Dateien abgerufen. Dies deutet auf Benutzer hin, die sich nicht lange auf den Seiten umsehen.

Monatsdurchschnitt		
<i>Abrufe ca.</i>	<i>Besuche ca.</i>	<i>Verhältnis</i>
3 500	1 400	2,50 : 1

Tagesdurchschnitt	
<i>Abrufe ca.</i>	<i>Besuche ca.</i>
120	50

4.2 Wochenverlauf

Die Nutzung der Seiten relativ zu Werktagen und Wochenende zeigt, dass die Hauptnutzung mit knapp drei Viertel unter der Woche erfolgt, mit knapp einem Viertel am Wochenende: auch hier spielt sich die Benutzung vorrangig während der Arbeitszeit ab.

<i>Wochentage</i>	<i>Anteil ca.</i>
Mo. – Fr.	78 %
Sa. – So.	22 %

4.3 Tagesverlauf

Über den Tag verteilt sich die Nutzung auf die üblichen drei Zeitfenster. Der Anteil an Nachtnutzern ist hier auffallend höher als in der Gesamtstatistik.

<i>Zeitfenster</i>	<i>Anteil ca.</i>
00:00 – 08:00 Uhr	11 %
08:00 – 19:00 Uhr	70 %
19:00 – 00:00 Uhr	19 %

4.4 Top-URLs

In der Liste der anteilmäßig am häufigsten abgerufenen Angebote führen deutlich Abfragen der eigentlichen Datenbank. Zu bemerken ist, dass davon drei Viertel auf die Kurztitelanzeigen entfallen und nur ein Viertel auf die vollständigen Anzeigen der Titel.

<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	74 % 55,5 % 18,5 %	Datenbank <i>Kurztitelanzeige</i> <i>Volltitelanzeige</i>
2	18 %	Einstiegsseite
3	4 %	Bodenseekarte
4	2 %	Sachgruppen
...		

4.5 Ein- und Ausgangsseiten

Der Beginn der Besuche liegt mit mehr als der Hälfte bei Datenbankabfragen. Da diese nicht von sonstigen Bereichen des Web-Angebotes verlinkt sind, stehen dahinter externe Besucher, die z. B. von den Euregio-Seiten des BSZ (<http://www.bsz-bw.de/eu/bodensee.html>) oder sonstigen externen Links einsteigen.

Dass 5 % Besuche mehr mit einer Datenbankabfrage enden, lässt sich dadurch erklären, dass die Besucher zwar mit einer anderen Seite beginnen (Einstiegsseite, Bodenseekarte, ...) und den Besuch mit einer Recherche beenden.

Eingangsseiten		
<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	57 %	Datenbank
2	18 %	Einstiegsseite
3	4 %	Bodenseekarte
...		

Ausgangsseiten		
<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	63 %	Datenbank
2	18 %	Einstiegsseite
3	7 %	Bodenseekarte
...		

4.6 Nutzerkreise

Die Verteilung auf Universitäts- und externe Nutzer gestaltet sich wie folgt:

<i>Nutzerkreis</i>	<i>Anteil ca.</i>
uni-konstanz.de	25 %
<i>davon UB-Mitarbeiter</i>	14 %
<i>davon UB-Nutzer</i>	6 %
<i>davon sonstige Uni</i>	5 %
extern	75 %

5 Auswertung „Datenbankangebot“

Die Auswertung „Datenbankangebot“ umfasst sowohl das Angebot der hauseigenen CD-ROM-Datenbanken, als auch das bei ReDI aufliegende Angebot.

Bislang nicht erfasst sind Zugriffe auf die Hilfedateien zu den jeweiligen Datenbanken, was allerdings in den Auswertungen ab Februar erfolgen wird.

5.1 Generelle Durchschnittswerte

Über die Monate betrachtet werden im Durchschnitt pro Besuch 6 Dateien abgerufen, im Vergleich zu den anderen Verhältnissen recht viel. Hier rufen die Benutzer also pro Besuch überdurchschnittlich viele Dateien ab.

Monatsdurchschnitt		
<i>Abrufe ca.</i>	<i>Besuche ca.</i>	<i>Verhältnis</i>
10 800	1 800	6,00 : 1

Tagesdurchschnitt	
<i>Abrufe ca.</i>	<i>Besuche ca.</i>
370	60

5.2 Wochenverlauf

In der Nutzung der Seiten relativ zu Werktagen und Wochenende zeigt eine höhere Konzentration der Abrufe auf die Werktage als auf das Wochenende.

<i>Wochentage</i>	<i>Anteil ca.</i>
Mo. – Fr.	84 %
Sa. – So.	16 %

5.3 Tagesverlauf

Über den Tag verteilt sich die Nutzung auf die üblichen drei Zeitfenster und deckt sich grob gesehen mit der Nutzung des gesamten Web-Angebotes.

<i>Zeitfenster</i>	<i>Anteil ca.</i>
00:00 – 08:00 Uhr	4 %
08:00 – 19:00 Uhr	84 %
19:00 – 00:00 Uhr	12 %

5.4 Top-URLs

Die Liste der anteilmäßig am häufigsten abgerufenen Angebote wird *in der Summe* von den statischen Einzelseiten angeführt: der Einstiegs- und Hilfeseite, den Fachlisten und der alphabetischen Liste.

Die Nutzung der Datenbanken kann auf Web-Server-Seite nur durch die Aufrufe der zur Nutzung notwendigen Plug-Ins (bei lokalen Datenbanken auf dem CD-ROM-Server) oder die „Sprungdateien“ (bei ReDI-Datenbanken) protokolliert werden, was quasi dem Aufruf der Datenbank entspricht; deren eigentliche Nutzung kann hierüber nicht ermittelt werden.

<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	65 %	Einzelne statische Seiten
	25 %	<i>Einstiegsseite</i>
	10 %	<i>Fachliste Sozialwiss.</i>
	10 %	<i>Fachliste Geisteswiss.</i>
	8 %	<i>alfabetische Liste</i>
	7 %	<i>Fachliste Naturwiss.</i>
	5 %	<i>Fachliste Allgemein</i>
2	21 %	ReDI-Datenbanken
3	14 %	CD-ROM-Server

5.5 Ein- und Ausgangsseiten

Der überwiegende Teil der Besuche startet mit der Einstiegsseite des Datenbankangebotes (*/daba/daba.htm*), kommt also den üblichen Weg über die Homepage der Bibliothek. Die Einstiege auf den Datenbanklisten rühren mit großer Wahrscheinlichkeit von den Links auf den Fachinformationsseiten her.

Die Zahlen der Ausgangsseiten sind hier nicht relevant, da sich das eigentliche Angebot ausserhalb des Web-Servers befindet, und ein Schritt der Besucher dorthin bereits als Ausgang gezählt wird.

Eingangsseiten		
<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	55 %	Einstiegsseite
2	13 %	Fachliste Geisteswiss.
3	10 %	alfabetische Liste
4	7 %	Fachliste Sozialwiss.
5	5 %	Fachliste Naturwiss.
6	3 %	Fachliste Allgemein
...		

5.6 Nutzerkreise

In der Verteilung auf Universitäts- und externe Nutzer treten immerhin 26 % externe Nutzer auf, die auch tatsächlich Inhalte geliefert bekommen. Dies dürfte sich vor allem auf die Lektüre der Datenbanklisten beschränken.

<i>Nutzerkreis</i>	<i>Anteil ca.</i>
uni-konstanz.de	74 %
<i>davon UB-Mitarbeiter</i>	10 %
<i>davon UB-Nutzer</i>	38 %
<i>davon sonstige Uni</i>	26 %
extern	26 %

6 Auswertung „englischsprachiges Angebot“

Diese Auswertung umfasst das Angebot der englischsprachigen Seiten.

6.1 Generelle Durchschnittswerte

Das über die Monate auftretende Verhältnis von Abrufen zu Besuchen von durchschnittlich 2,89 abgerufenen Dateien pro Besuch deckt sich gut mit dem Durchschnitt.

Monatsdurchschnitt		
<i>Abrufe ca.</i>	<i>Besuche ca.</i>	<i>Verhältnis</i>
1 300	450	2,89 : 1

Tagesdurchschnitt	
<i>Abrufe ca.</i>	<i>Besuche ca.</i>
44	14

6.2 Wochenverlauf

In der Nutzung der Seiten relativ zu Werktagen und Wochenende tritt wieder eine Konzentration der Nutzung auf die Werktage hervor.

<i>Wochentage</i>	<i>Anteil ca.</i>
Mo. – Fr.	83 %
Sa. – So.	17 %

6.3 Tagesverlauf

Über den Tag verteilt sich die Nutzung auf die üblichen drei Zeitfenster. Anteilsmäßig ist die Nutzung zwischen Mitternacht und 08:00 Uhr mit einem Viertel recht hoch.

<i>Zeitfenster</i>	<i>Anteil ca.</i>
00:00 – 08:00 Uhr	25 %
08:00 – 19:00 Uhr	56 %
19:00 – 00:00 Uhr	19 %

6.4 Top-URLs

Die Liste der anteilmäßig am häufigsten abgerufenen Angebote lässt neben der Einstiegsseite ein überwiegendes Interesse an der Seite mit den deutschen und internationalen Katalogen, den elektronischen Publikationen und den Datenbanken erkennen.

<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	23 %	Einstiegsseite
2	5 %	Kataloge.htm
3	3 %	epubli.htm
4	2 %	Daba.htm
5	2 %	search.htm
6	2 %	Links.htm
...		

6.5 Ein- und Ausgangsseiten

Interessant an den Startseiten der Besuche ist das recht rege Interesse an der Fahrplanauskunft und vor allem an der Liste der Kataloge.

Ein Wert von 42 % der Einstiegsseite (`/eng/index.htm`) als Ausgangsseite deutet darauf hin, dass dies solche „Besuche“ sind, bei denen der Nutzer nur einen kurzen Blick auf die Seite wirft.

Eingangsseiten		
<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	64 %	Einstiegsseite
2	3 %	Kataloge.htm
3	2 %	Fahrplanauskunft
...		

Ausgangsseiten		
<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	42 %	Einstiegsseite
2	9 %	Kataloge.htm
3	5 %	epubli..
...		

6.6 Nutzerkreise

Die Verteilung auf Universitäts- und externe Nutzer weist immerhin einen recht vorzeigbaren Anteil an internen Nutzern auf, von denen die Mitarbeiter mit den geringsten Teil ausmachen.

<i>Nutzerkreis</i>	<i>Anteil ca.</i>
uni-konstanz.de	33 %
<i>davon UB-Mitarbeiter</i>	5 %
<i>davon UB-Nutzer</i>	20 %
<i>davon sonstige Uni</i>	8 %
extern	67 %

7 Auswertung „Fachinformationsseiten“

Die Auswertung „Fachinformationsseiten“ umfasst die Seiten der Fachreferenten. Hier wurden, im Gegensatz zu den anderen Auswertungen, nicht einzelne Seiten ausgewertet, sondern deren Summe unter dem Blickwinkel des jeweiligen Faches. Sämtliche Zugriffe sind daher auf der zweiten Ebene zusammengefasst.

7.1 Generelle Durchschnittswerte

Über die Monate betrachtet werden im Durchschnitt 2,61 Dateien pro Besuch abgerufen, was auch hier auf schnellen und kurzen Zugriff auf die Seiten hindeutet.

Monatsdurchschnitt		
<i>Abrufe ca.</i>	<i>Besuche ca.</i>	<i>Verhältnis</i>
31 100	11 900	2,61

Tagesdurchschnitt	
<i>Abrufe ca.</i>	<i>Besuche ca.</i>
1 000	390

7.2 Wochenverlauf

Die Nutzung der Seiten relativ zu Werktagen und Wochenende weist hier keine besondere Verteilung auf, sondern deckt sich mit der Nutzung des kompletten Web-Angebotes.

<i>Wochentage</i>	<i>Anteil ca.</i>
Mo. – Fr.	84 %
Sa. – So.	16 %

7.3 Tagesverlauf

In der Verteilung der Nutzung über den Tag hinweg war bei dieser Auswertung ein viertes Zeitfenster erforderlich, da der Durchlauf des Volltextindexierungsprogrammes „*ht://dig*“ um 08:30 die Nutzung im entsprechenden Zeitfenster überdeutlich ansteigen lässt.

Bezieht man den sonst vorherrschenden Wert auf dem vorigen und anschließenden Zeitsegment von mit ein, so beträgt der Anteil dieser Suchmaschine rund 25 %.

In den Auswertungen ab Februar werden auch Zugriffe der Suchmaschine nicht beachtet, so dass hier ein unverfälschter Wert ersichtlich sein wird.

Die Verteilung an sich ist, verglichen mit den sonstigen Werten, hier jedoch recht flach.

<i>Zeitfenster</i>	<i>Anteil ca.</i>
00:00 – 08:00 Uhr	11 %
08:00 – 09:00 Uhr	28 %
09:00 – 19:00 Uhr	45 %
19:00 – 00:00 Uhr	16 %

7.4 Top-URLs

In der Liste der anteilmäßig am meisten abgerufenen Angebote verteilen sich die ersten fünf Fächer wie folgt:

<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	18 %	pha
2	10 %	jura
3	10 %	edz
4	8 %	rom
5	6 %	arc
...		

7.5 Ein- und Ausgangsseiten

Die Startseiten der Besuche liegen überwiegend in den einzelnen Angeboten: die Besucher kommen also direkt auf diese Seiten: entweder von externen Angeboten oder über Bookmarks.

Eingangsseiten		
<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	18 %	pha
2	9 %	jura
3	9 %	ant
4	8 %	rom
5	8 %	arc
...		

Ausgangsseiten		
<i>Position</i>	<i>Anteil ca.</i>	<i>Angebot</i>
1	19 %	pha
2	10 %	jura
3	9 %	ant
4	9 %	rom
5	8 %	arc
...		

7.6 Nutzerkreise

In Verteilung der Nutzung auf verschiedene Nutzerkreise ist der überaus hohe Anteil an externen Benutzern auffallend.

<i>Nutzerkreis</i>	<i>Anteil ca.</i>
uni-konstanz.de	36 %
davon UB-Mitarbeiter	31 %
davon UB-Nutzer	3 %
davon sonstige Uni	2 %
extern	64 %

8 Fazit und Ausblick

8.1 Zum Ergebnis

Als Ergebnis des Projektes liegen jetzt einige Zahlen vor, die die Nutzung des Web-Angebotes der Bibliothek recht deutlich darstellen und teils auch genaue Auskunft über die Zusammensetzung der Nutzerschaft, wie auch über die Herkunft der Besuche geben.

Eine weiter gehende Interpretation dieser Zahlen krankt vor allem an der Tatsache, dass keine Vergleichszahlen vorliegen.

Am Beispiel der Nutzung der Fachinformationsseiten lässt sich dies demonstrieren: ist der Abruf von 2,61 Seiten pro Besuch nun gut oder schlecht? Erklärungsvariante Eins wäre: es ist gut, denn die Nutzer finden schnell die gewünschte Information. Erklärungsvariante Zwei wäre: es ist schlecht, denn die Nutzer schöpfen nicht das volle Potential der Seiten aus.

Als zufrieden stellend könnte die Tatsache betrachtet werden, dass die Fachinformations-Seiten, zwar mit nur 5 % Anteil, immerhin den zweit stärksten Anteil an der Gesamtnutzung ausmachen. In Frage gestellt werden muss dagegen der hohe Anteil an externen Nutzern: im Grunde legt dies den Schluss nahe, dass der Hauptteil der Arbeit *nicht* für die eigene Kundschaft, also die Benutzer der Universität erfolgt. Jedoch müsste hier in Betracht gezogen werden, wie viele dieser externen Nutzer heimschaffende Studierende, bzw. sonstige Universitätsangehörige sind, was außerhalb des Bereich des Möglichen liegt.

Doch auch unter dem Blickwinkel der Gesamtnutzung, die mit 71 % von der eigentlichen Kundschaft, den Universitätsangehörigen, und hiervon sogar mit 46 % in den Buchbereichen selbst erfolgt, sind die Ergebnisse positiv zu deuten, allerdings mit der Einschränkung, dass der Hauptanteil hier auf die Katalogabfrage entfällt.

Um genauere Aussagen treffen zu können wäre es zudem notwendig zu erheben, in welchem Maße die Nutzer auf die elektronischen Angebote „ausweichen“: hier ist durchaus zu berücksichtigen, dass ein brauchbarer Informationsbestand in konventioneller Form, sprich: Bücher, Nachschlagewerke, etc., vor allem aber auch gedruckt angebotene Fachführer, die Nutzung der Web-Angebote erübrigt, und natürlich umgekehrt.

8.2 Fortführung

Die tägliche Auswertung erfolgt nach gegebenem Muster. Für die Auswertung über einen definierten Zeitraum hinweg ist es jedoch notwendig, die relevanten Daten monatlich für entsprechende Programme zusätzlich zu erfassen, z. B. durch Eingabe in eine Tabellenkalkulation. Dieser Schritt wurde bei diesem Vorgehen versäumt.

Möglich wären damit vor allem differenzierte Untersuchungen für eine von der täglichen Auswertung unabhängige und teils auch weitergehende Statistik, die eher als der tägliche automatisierte Vorgang die Beantwortung von differenzierten Einzelfragen ermöglicht.

9 Anhänge

9.1 Software

Als Software wird verwendet:

- Webalizer, V. 2.01 (<http://www.mrunix.net/webalizer/>)
- GD Graphics Library (<http://www.boutell.com/gd/>)
- Berkeley DB Library (<http://www.sleepycat.com/>)
- eigens verfasste Perl-Skripte

9.2 Anpassung

Um Auswertungen für neue Teilbereiche einzurichten ist erforderlich:

1. Anlegen einer Konfigurationsdatei unter `/usr/local/webalizer/conf` nach dem Muster `name.conf`
2. Aufnehmen des Schlüssels `name` in die Liste im Skript `/usr/local/webalizer/auswertung.pl`
3. Anlegen eines Ausgabeverzeichnis unter `/u/www/htm/intranet/webalizer/`

9.3 Datei `parse.pl`

Dieses Skript wird aufgerufen, um die Datenmanipulationen durchzuführen.

```
#!/usr/bin/perl -w

use strict;

open(IN,'< /usr/local/apache/var/log/access_log');
open(KNUB,'> /tmp/knub.log');
open(FI,'> /tmp/fi.log');
while(<IN>) {
# wir zaehlen keine Grafik-Dateien
  next if /\.gif HTTP\/1\./o;
  next if /\.jpg HTTP\/1\./o;
  next if /\.jpeg HTTP\/1\./o;
  next if /\.css HTTP\/1\./o;
  next if /\.js HTTP\/1\./o;
  next if /\/_derived\/o;
#
#Zugriffe auf Web-Root ignorieren
  next if /"GET \/ HTTP\/1\./o;
# Zugriffe auf Web-Root ignorieren
  next if /"GET \/index\.htm HTTP\/1\./o;
# Zugriffe auf robot-Datei ignorieren
  next if /T \/robots\.txt HTTP\/1\./o;
# HEAD-Abfragen rauswerfen
```

```

next if /"HEAD \/\/o;
# 0-Byte-Antworten.
next if /HTTP\/1\.[01]" 200 [-0]$/o;
# Zugriffe von w1 ignorieren
next if /^134\.34\.34\.9/o;
# Zugriffe von w3 ignorieren
next if /^134\.34\.200\.235/o;
# fingierte Hostnamen
s/^134\.34\.200\.[0-9]{1,3}/mitarbeiter.ub.uni-konstanz.de/o;
s/^134\.34\.34\.[0-9]{1,3}/benutzer.ub.uni-konstanz.de/o;
s/^134\.34\.[0-9]{1,3}\.[0-9]{1,3}/sonstige.uni-konstanz.de/o;
# nur noch mit 200-Code weitermachen
speziell(\$_) if /HTTP\/1\.[01]" 200 /o;
}
close(FI);
close(KNUB);
close(IN);

sub speziell {
    my $line = shift;
    print KNUB $$line; # speichern
# Fachinfo wird gesondert behandelt
    print FI if $$line =~ s/T \\/fi\/(\w{2,4})\/.*
                HTTP\/1/T \\/fi\/$1 HTTP\/1/o;
    print FI if $$line =~ /T \\/Fachinfo\.htm HTTP\/1\./o;
}

```

9.4 Datei auswertung.pl

Dieses Skript startet die verschiedenen Durchläufe des Webalizer-Programmes.

```

#! /usr/bin/perl

use strict;

my $items = ['allg', 'daba', 'doklfg', 'eng', 'epubli', 'fi',
             'kataloge', 'regio', 'bodbibl', 'sis', 'semcgi',
             'intra'];

my $start = scalar localtime;

print "\n*** Resolving DNS-Addresses ***\n\n";
system('./webazolver -D
        /u/www/htm/intranet/webalizer/dns_cache.db
        -N 20 /tmp/knub.log');

foreach (@{$items}) {
    print "\n *** Running '$_' ***\n\n";
    system("./webalizer -c ./conf/$_conf");
}

```



```

my $end = scalar localtime;

print "\n *** DONE :-)\n\n";
print " => Start: $start\n => Ende: $end\n";

```

9.5 Konfigurationsdatei allg.conf

```

LogFile          /tmp/knub.log
OutputDir        /u/www/htm/intranet/webalizer/allg/
ReportTitle      Web allgemein /
TimeMe           yes
Incremental      yes
HostName         www.ub.uni-konstanz.de
TopSites         100
TopURLs          100
TopKURLs         100
TopEntry         25
TopExit          25

DNSCache         /u/www/htm/intranet/webalizer/dns_cache.db
DNSChildren      0

GroupDomains     1

GroupURL /fi/*
GroupURL /daba/*

HourlyStats      yes
DailyStats       yes
GraphLegend      yes
DailyGraph       yes
HourlyGraph      yes

```

9.6 Konfigurationsdatei bodbibl.conf

```

LogFile          /tmp/knub.log
OutputDir        /u/www/htm/intranet/webalizer/bodbibl/
ReportTitle      BodBibl /
TimeMe           yes
Incremental      yes
HostName         www.ub.uni-konstanz.de
TopSites         100
TopURLs          100
TopKURLs         100
TopEntry         25
TopExit          25

DNSCache         /u/www/htm/intranet/webalizer/dns_cache.db
DNSChildren      0

```

```

HourlyStats      yes
DailyStats       yes
GraphLegend      yes
DailyGraph       yes
HourlyGraph      yes

GroupDomains     1

PageType         pl
PageType         htm
PageType         html

IgnoreURL        *
IncludeURL       /cgi-bin/acwww25/allegro_bod.pl
IncludeURL       /cgi-bin/acwww25/regsrch_bod.pl
IncludeURL       /boddb/*

```

9.7 Konfigurationsdatei daba.conf

```

LogFile          /tmp/knub.log
OutputDir        /u/www/htm/intranet/webalizer/daba/
ReportTitle      Datenbankangebot /
TimeMe           yes
Incremental      yes
HostName         www.ub.uni-konstanz.de
TopSites         50
TopURLs          100
TopEntry         25
TopExit          25

DNSCache         /u/www/htm/intranet/webalizer/dns_cache.db
DNSChildren      0

HourlyStats      yes
DailyStats       yes
GraphLegend      yes
DailyGraph       yes
HourlyGraph      yes

GroupDomains     1
GroupURL         *.ica
GroupURL         /_redi/*
GroupURL         /_cdrom2/*

IgnoreURL        *
IncludeURL       /daba/*
IncludeURL       /_redi/*
IncludeURL       /ica/*
IncludeURL       /_cdrom2/*

```

9.8 Konfigurationsdatei doklfg.conf

```
LogFile          /tmp/knub.log
OutputDir        /u/www/htm/intranet/webalizer/doklfg/
ReportTitle      Dokumentenlieferung /
TimeMe           yes
Incremental      yes
HostName         www.ub.uni-konstanz.de
TopSites         100
TopURLs          100
TopEntry         25
TopExit          25

DNSCache         /u/www/htm/intranet/webalizer/dns_cache.db
DNSChildren      0

HourlyStats      yes
DailyStats       yes
GraphLegend      yes
DailyGraph       yes
HourlyGraph      yes

GroupDomains     1

IgnoreURL        *
IncludeURL        /Dokulief.htm
IncludeURL        /subito03.htm
IncludeURL        /cgi-bin/flwbest
IncludeURL        /cgi-bin/tb/dod*
IncludeURL        /cgi-bin/subito.pl
IncludeURL        /cgi-bin/acwww25/subitocgi.pl
```

9.9 Konfigurationsdatei eng.conf

```
LogFile          /tmp/knub.log
OutputDir        /u/www/htm/intranet/webalizer/eng/
ReportTitle      Englischsprachiges Angebot /
TimeMe           yes
Incremental      yes
HostName         www.ub.uni-konstanz.de
TopSites         100
TopURLs          100
TopEntry         25
TopExit          25

DNSCache         /u/www/htm/intranet/webalizer/dns_cache.db
DNSChildren      0

HourlyStats      yes
DailyStats       yes
```

```

GraphLegend    yes
DailyGraph     yes
HourlyGraph    yes

GroupDomains   1

IgnoreURL      *
IncludeURL     /eng/*

```

9.10 Konfigurationsdatei epubli.conf

```

LogFile        /tmp/knub.log
OutputDir      /u/www/htm/intranet/webalizer/epubli/
ReportTitle    KOPS / Epubli /
TimeMe        yes
Incremental    yes
HostName       www.ub.uni-konstanz.de
TopSites       100
TopURLs        300
TopKURLs       300
TopEntry       25
TopExit        25

DNSCache       /u/www/htm/intranet/webalizer/dns_cache.db
DNSChildren    0

HourlyStats    yes
DailyStats     yes
GraphLegend    yes
DailyGraph     yes
HourlyGraph    yes

GroupDomains   1
GroupURL       /esem/*

IgnoreURL      *
IncludeURL     /kops/*
IncludeURL     /v13/*
IncludeURL     /cgi-bin/w3-msql/*
IncludeURL     /cgi/w3-msql/*
IncludeURL     /epubli.htm
IncludeURL     /ebooks/*
IncludeURL     /esem/*

```

9.11 Konfigurationsdatei fi.conf

```

LogFile        /tmp/fi.log
OutputDir      /u/www/htm/intranet/webalizer/fi/
ReportTitle    Fachinformation /
Debug          no

```

```

TimeMe          yes
Incremental     yes
HostName        www.ub.uni-konstanz.de
TopSites        100
TopURLs         100
TopKURLs        100
TopEntry        25
TopExit         25

DNSCache        /u/www/htm/intranet/webalizer/dns_cache.db
DNSChildren     0

HourlyStats     yes
DailyStats      yes
GraphLegend     yes
DailyGraph      yes
HourlyGraph     yes

GroupDomains    1

IgnoreURL       *
IncludeURL      /fi/
IncludeURL      /Fachinfo.htm

```

9.12 Konfigurationsdatei intra.conf

```

LogFile         /tmp/knub.log
OutputDir       /u/www/htm/intranet/webalizer/intra/
ReportTitle     Intranet /
TimeMe          yes
Incremental     yes
HostName        www.ub.uni-konstanz.de
TopSites        100
TopURLs         100
TopKURLs        100
TopEntry        25
TopExit         25

DNSCache        /u/www/htm/intranet/webalizer/dns_cache.db
DNSChildren     0

HourlyStats     yes
DailyStats      yes
GraphLegend     yes
DailyGraph      yes
HourlyGraph     yes

GroupDomains    1

IgnoreURL       *

```

```
IncludeURL      /_intranet/*
IncludeURL      /intranet/*
# Datei nicht zugriffsgeschuetzt
#IncludeURL     /intranet.htm
```

9.13 Konfigurationsdatei kataloge.conf

```
LogFile         /tmp/knub.log
OutputDir       /u/www/htm/intranet/webalizer/kataloge/
ReportTitle     Kataloge /
TimeMe         yes
Incremental     yes
HostName        www.ub.uni-konstanz.de
TopSites        100
TopURLs         100
TopKURLs        100
TopEntry        25
TopExit         25

DNSCache        /u/www/htm/intranet/webalizer/dns_cache.db
DNSChildren     0

HourlyStats     yes
DailyStats      yes
GraphLegend     yes
DailyGraph      yes
HourlyGraph     yes

GroupDomains    1

IgnoreURL       *
IncludeURL      /cgi-bin/acwww25/allegro.pl
IncludeURL      /cgi-bin/acwww25/regsrch.pl
IncludeURL      /cgi-bin/acwww25/acindex2.pl
IncludeURL      /cgi-bin/acwww25/sig.pl
IncludeURL      /cgi-bin/acwww25/asitcgi.pl
IncludeURL      /cgi-bin/kontoinf.pl
IncludeURL      /koala/*
```

9.14 Konfigurationsdatei regio.conf

```
LogFile         /tmp/knub.log
OutputDir       /u/www/htm/intranet/webalizer/regio/
ReportTitle     RegioKat / BodBibl /
TimeMe         yes
Incremental     yes
HostName        www.ub.uni-konstanz.de
TopSites        100
TopURLs         100
```

```

TopKURLs      100
TopEntry      25
TopExit       25

DNSCache      /u/www/htm/intranet/webalizer/dns_cache.db
DNSChildren   0

HourlyStats   yes
DailyStats    yes
GraphLegend   yes
DailyGraph    yes
HourlyGraph   yes

GroupDomains  1

PageType      pl
PageType      htm
PageType      html

IgnoreURL     *
IncludeURL    /cgi-bin/acwww25/allegro_regio.pl
IncludeURL    /cgi-bin/acwww25/regsrch_regio.pl
IncludeURL    /regio/*
IncludeURL    /Euregio.htm
IncludeURL    /euregio

```

9.15 Konfigurationsdatei semcgi.conf

```

LogFile       /tmp/semcgi.log
OutputDir     /u/www/htm/intranet/webalizer/semcgi/
ReportTitle   Sem-CGI /
TimeMe        yes
Incremental   yes
HostName      www.ub.uni-konstanz.de:8080
TopSites      100
TopURLs       100
TopKURLs      100
TopEntry      25
TopExit       25

DNSCache      /u/www/htm/intranet/webalizer/dns_cache.db
DNSChildren   0

HourlyStats   yes
DailyStats    yes
GraphLegend   yes
DailyGraph    yes
HourlyGraph   yes

GroupDomains  1

```

```
IgnoreURL      *
IncludeURL     /sem-cgi/sem/sem.pl
```

9.16 Konfigurationsdatei sis.conf

```
LogFile        /tmp/knub.log
OutputDir      /u/www/htm/intranet/webalizer/sis/
ReportTitle    SIS /
TimeMe         yes
Incremental    yes
HostName       www.ub.uni-konstanz.de
TopSites       100
TopURLs        100
TopKURLs       100
TopEntry       25
TopExit        25

DNSCache       /u/www/htm/intranet/webalizer/dns_cache.db
DNSChildren    0

HourlyStats    yes
DailyStats     yes
GraphLegend    yes
DailyGraph     yes
HourlyGraph    yes

GroupDomains   1

IgnoreURL      *
IncludeURL     /sis/
```