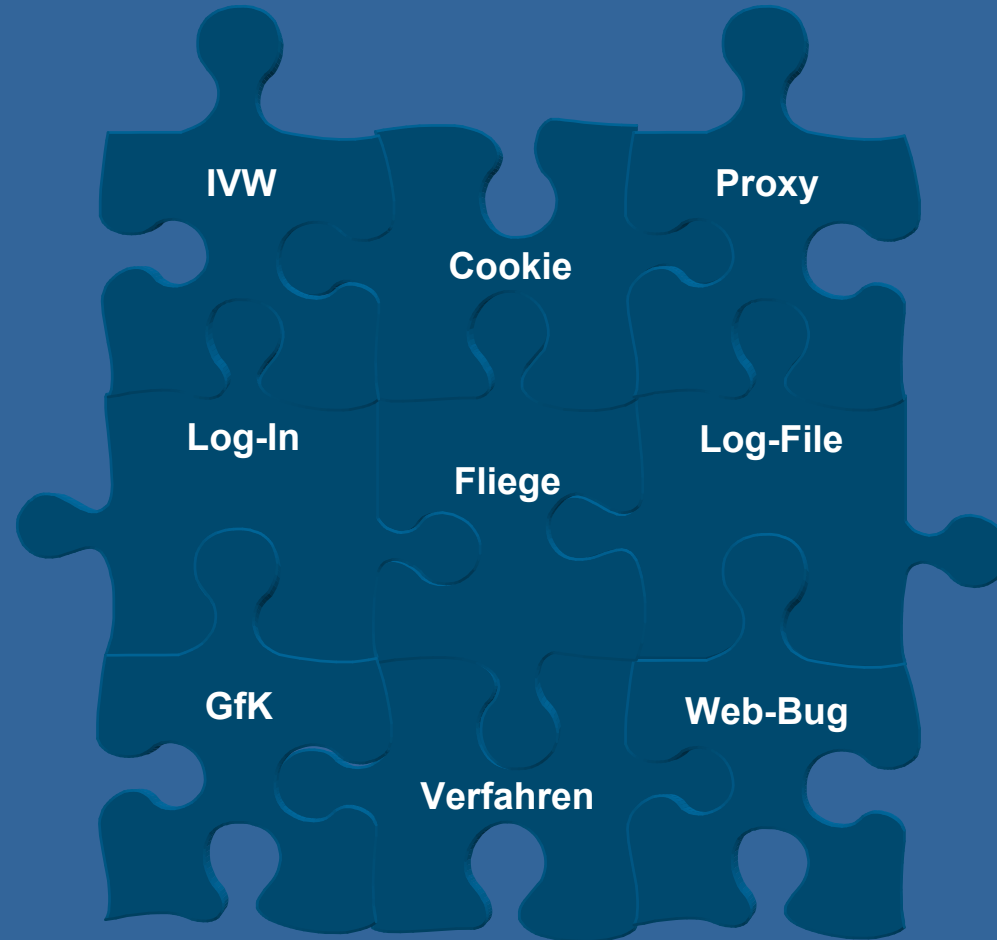
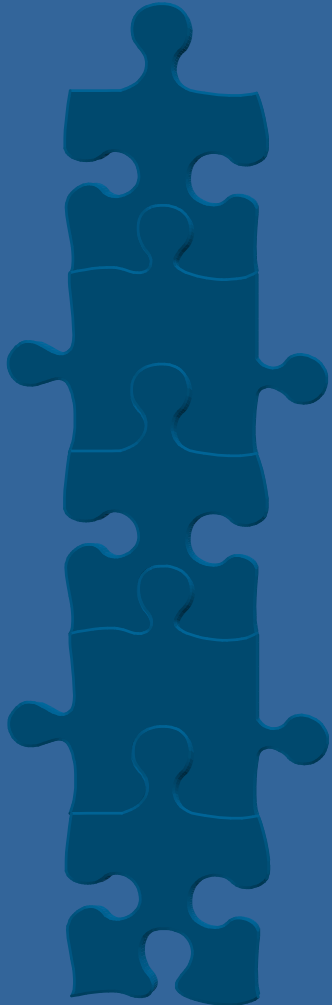


Marktforschung im Internet – Möglichkeiten der Log-File Analyse



Inhalt



1. Aufbau und Inhalt von Log-Files
2. Log-File Analyse
3. Problematik von Proxy-Servern
4. Neue Methoden der Leistungsmessung
5. Werbeträgerforschung durch Institute

Zielsetzung: Protokollierung der Seitenabrufe

- Fehleranalyse – technische Optimierung
- Attraktivität der Seiten – Optimierung der Web-Site
- Optimierung Werbemittel (Suchmaschineneintrag, Banner, ...)
- Werbeträgerleistung
- ...

1. Aufbau und Inhalt von Log-Files

- Pull Prinzip:
Nutzer rufen Online-Medien ab.
Damit sind zählbar und ggf. identifizierbar
- Daten werden automatisch in **Log-Files** gesammelt
- Bekannt sind Log-Files auch als „Spuren im Internet“ oder „elektronischer Fingerabdruck“

1. Aufbau und Inhalt von Log-Files

Beispiel eines Log-Files

194.175.62.68 weinberg - [01/Jun/2001:08:26:07 +0200]

"GET /produkte/gluehbirne.html HTTP/1.1" 200 78540

http://de.google.yahoo.com/bin/query_de?p=Gl%fcbirne&hc=0&hs=0

"Mozilla/4.0 (compatible; MSIE 5.5; Windows NT 4.0)" weinbergjxp4711

- Name oder IP-Adresse des Zugriffsrechners
- Name des Users (aber nur bei Login)
- Datum und Uhrzeit (und Abweichung zur Greenwich-Mean-Time GMT)
- Transferangabe, welche Seite wurde abgerufen
- Version des verwendeten Protokolls HTTP

1. Aufbau und Inhalt von Log-Files

Beispiel eines Log-Files

194.175.62.68 weinberg - [01/Jun/2001:08:26:07 +0200]

"GET /produkte/gluehbirne.html HTTP/1.1" 200 78540

http://de.google.yahoo.com/bin/query_de?p=Gl%fcbirne&hc=0&hs=0

"Mozilla/4.0 (compatible; MSIE 5.5; Windows NT 4.0)" weinbergjxp4711

- Status der Übertragung (200 = erfolgreich, 400 = abgebrochen, ...)
- Größe der übertragenen Datei
- Referrer = vorherige Seite
- Browserversion, Betriebssystem
- Inhalte von Cookies

1. Aufbau und Inhalt von Log-Files

Je nach Konfiguration werden nur ausgewählte oder alle möglichen Informationen ins Log-File geschrieben

Angaben im Log-File

- Browser
- Browserversion
- Betriebssystem
- IP-Adresse / Rechnername
- Login-Name
- Vorherige Seite (referrer)
- Gewünschte Seite (request)
- Datum und Zeitpunkt der Anfrage
- Inhalte von Cookies

1. Aufbau und Inhalt von Log-Files

Auswertung der Log-Files ergibt Auskünfte über das Nutzungsverhalten:

- Von welcher WWW-Seite kommen die Nutzer?
- Zu welcher Uhrzeit ist die Seite am höchsten frequentiert?
- Wie heißen die Nutzer?
- Welchen Browser benutzen sie?
- Welche Seite wird wie häufig abgerufen?
- Wie wird „navigiert“ (Click-Stream)

- Leistungskennziffern / Optimierungsansätze

1. Aufbau und Inhalt von Log-Files

„Klassische Leistungskennziffern:

- Page Impression
- Ad Impression
- View Time
- Visit
- Visit Time
- Click-Streams
- etc.

1. Aufbau und Inhalt von Log-Files

Problembereich: eingeschränkte Aussagekraft bei ‚naivem‘ Zählen:

- Mehrere User nutzen einen Rechner
- IP-Adressen sind oftmals dynamisch und wechseln
- Gateways vergeben eine Adresse für alle User
- Framework und Dummy-Grafiken erzeugen ebenfalls Log-Files
- Viele Log-Files pro Seite
- Proxy-Problematik
- etc.

1. Aufbau und Inhalt von Log-Files

Im kommerziellen Bereich werden neben „statische WWW-Seiten“ oft „dynamische Inhalte“ angeboten

→ Statische Seiten lassen sich leicht per Logfile Analyse auswerten

→ Dynamische Seiten erlauben besseres „Tracking“

Durch erweiterte Technik hat sich der Seitenaufbau verändert:

- Eine Web-Seite besteht aus einzelnen Objekten (Frames)
- Teilinhalte (z.B. Werbebanner) werden dynamisch geschaltet (rotierend)
- Teilinhalte (z.B. Banner) können individuell geschaltet werden
 - Ausrichtung auf gezielte Nutzergruppen
 - Anpassung an Interessen der Nutzer

2. Log-File Analyse

- Das automatische Auswerten von Log-Files erfolgt durch den Einsatz von sog. Counter-Software für Log-Server
- Für das Auswerten von privaten Servern gibt es zahlreiche Freeware und Shareware Angebote

Beispiel einer Counter-Software:

Log-Analyser von Surfe-Trace (www.surftrace.de)

- Auswertung von Inter- und Intranet-Logdateien in den Formaten Common, Combined, Agent und Referrer Log File
- Umwandlung der in einer Logdatei gespeicherten numerischen IP-Adressen in die entsprechenden Domain-Namen mittels Reverse DNS Lookup
- Erstellung von Ranglisten der beliebtesten Seiten, der aktivsten Top-Level-Domains, der am häufigsten verwendeten Browser und Betriebssysteme, ...
- Ermittlung der User-Sessions: Über welche Seiten gelangen die Besucher zum Angebot? Wie lange halten sie sich dort auf?

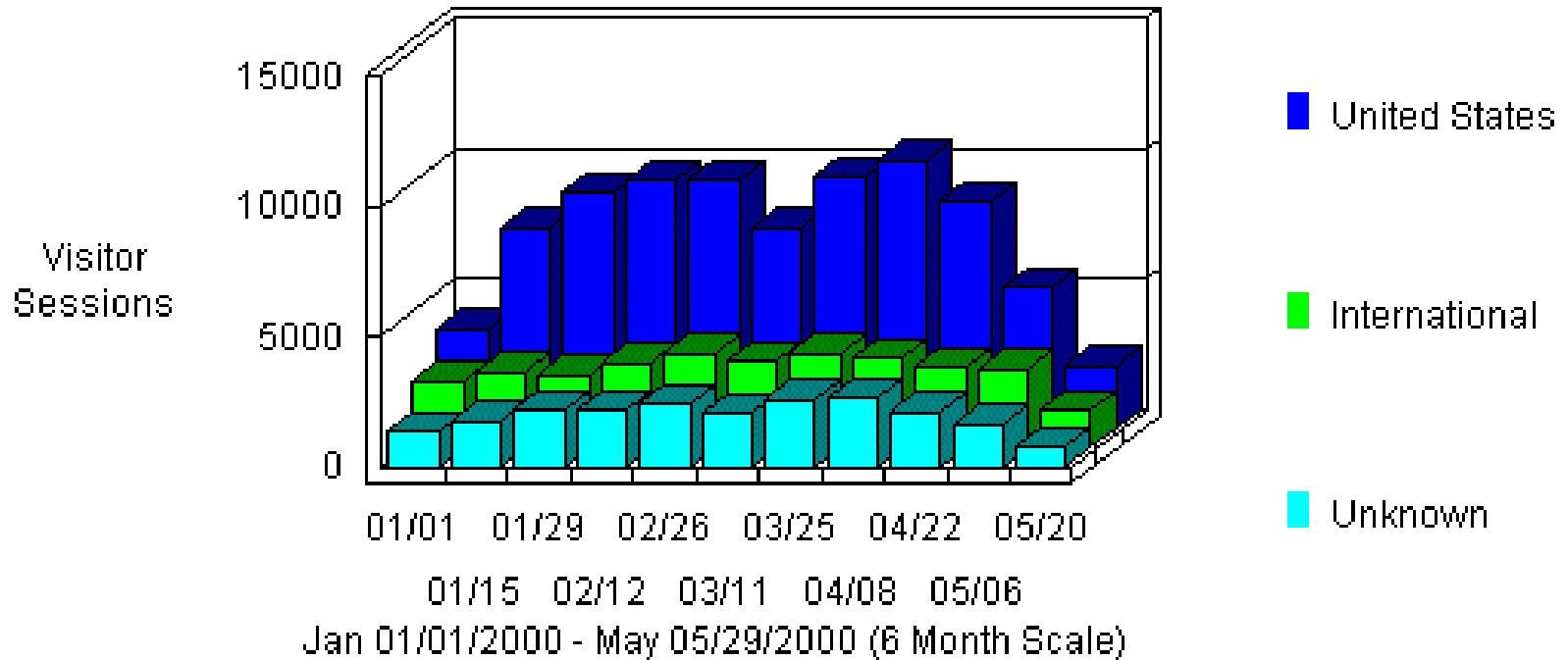
Auswertungsbeispiele

Einfache Zusammenfassung

Full Statistics for May 1999	
Monthly Summary	
Total hits	134801
Total files sent	115305
Total Code 304's (Not Modified)	15167
Other responses (see below)	4329
Total pageviews	46849
Remaining responses	87952
Total KB requested	1105079
Total KB transferred	983422
Total KB saved by cache	121657
Total unique URLs	1161
Total unique sites	12015
Total user sessions per 24 hours	17890
Total unique agents	1798
Total unique referrer URLs	9917
Total authenticated requests	1684
Maximum/Average hits	
Max hits per day	6602
Average hits per day	4348
Max hits per hour	1235
Average hits per hour	181

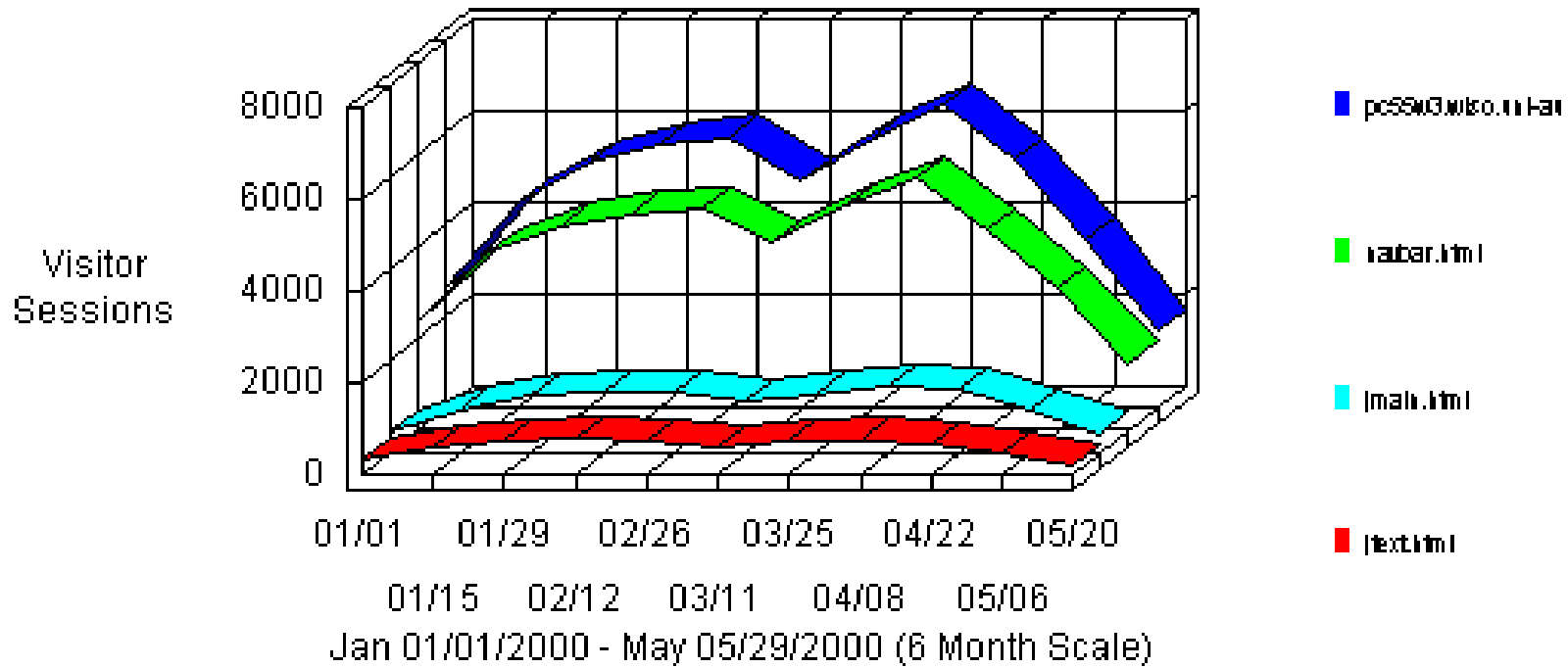
Auswertungsbeispiele

Visitor Profile by Region

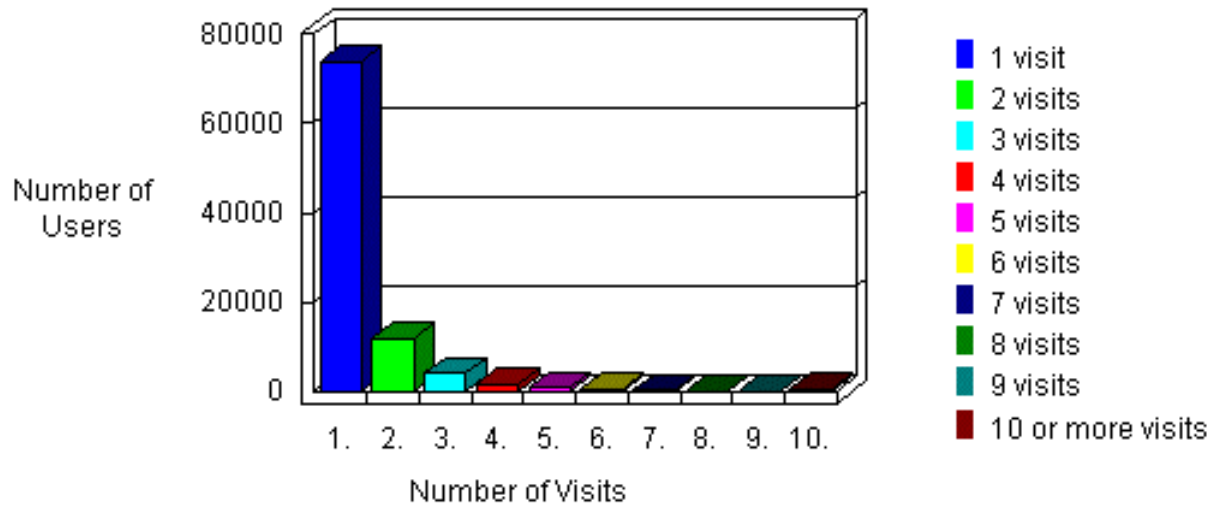


Auswertungsbeispiele

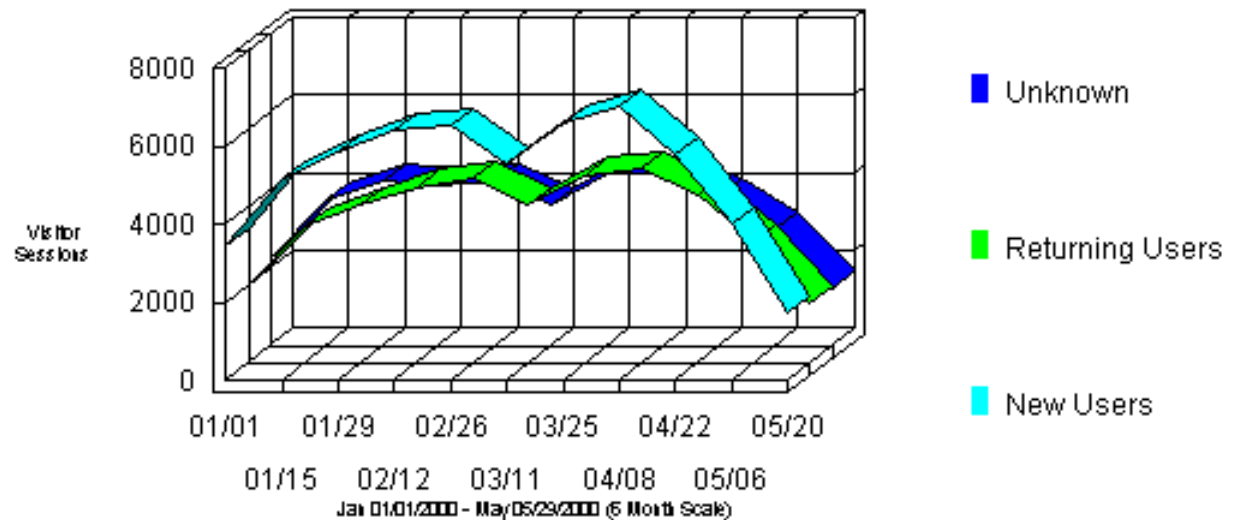
Most Requested Pages



Auswertungsbeispiele

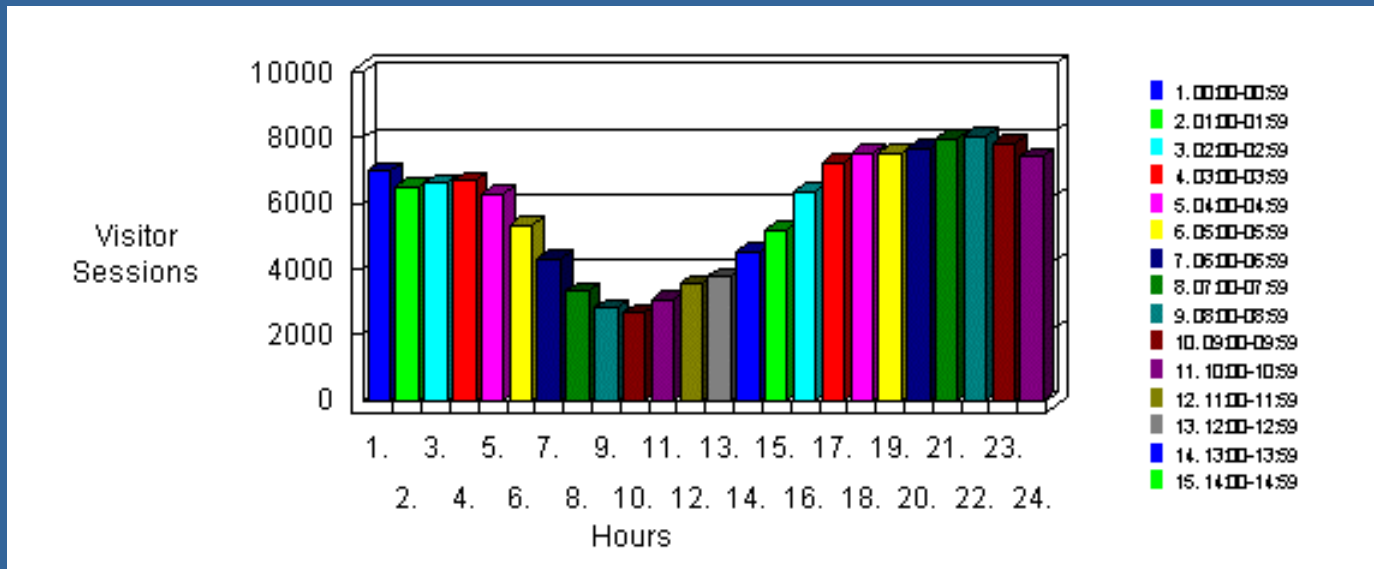
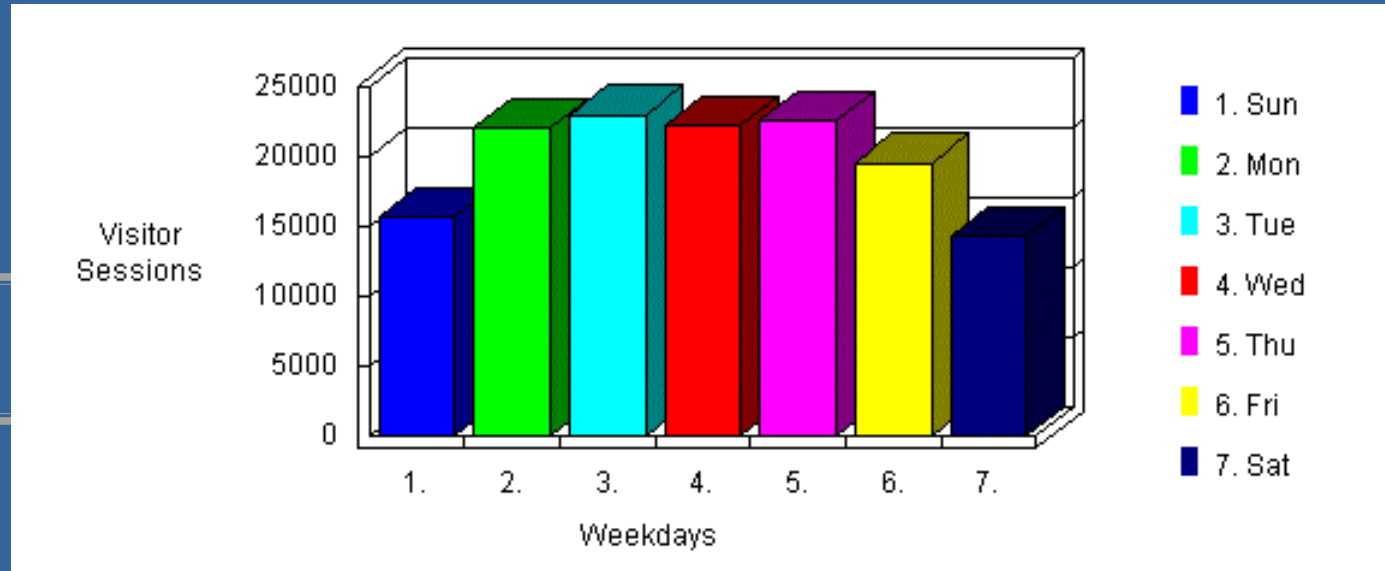


Besucher

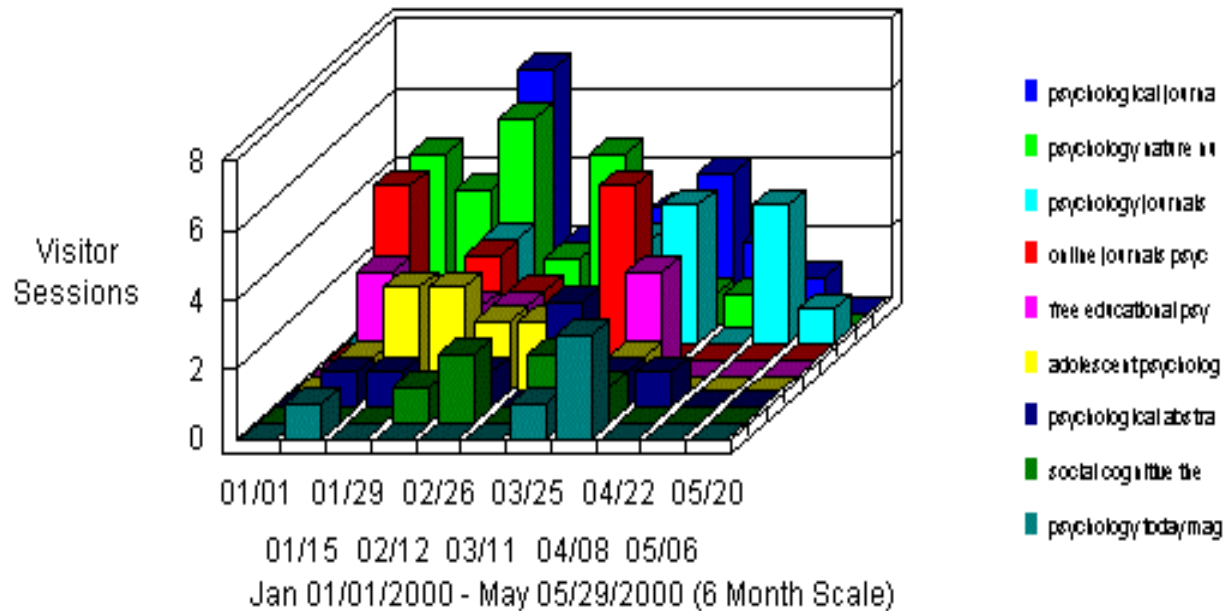


Auswertungsbeispiele

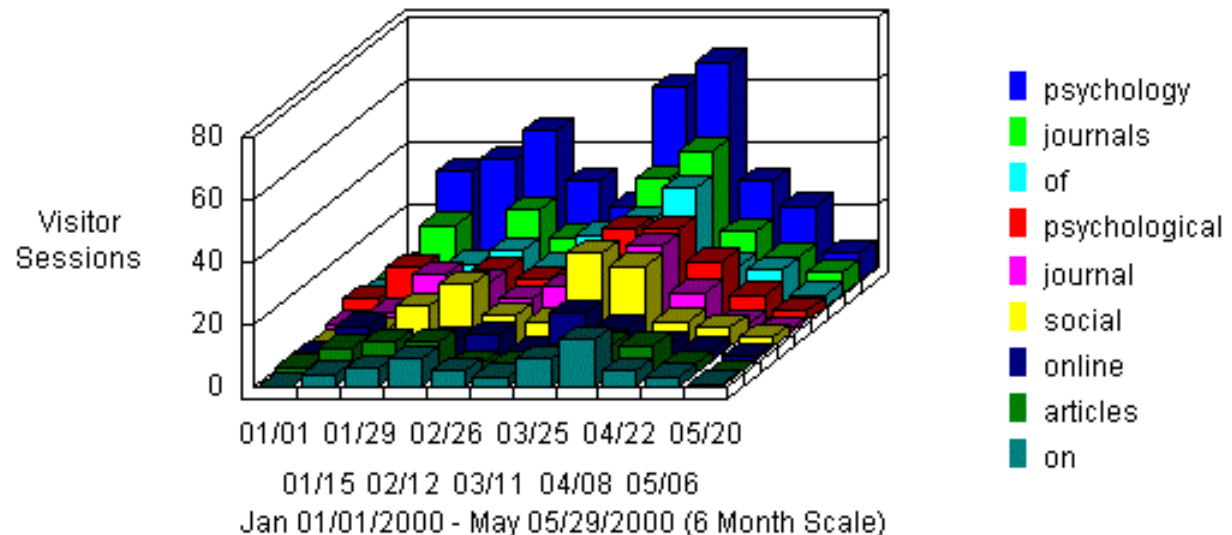
Zeitabhängigkeit



Auswertungsbeispiele



Suchbegriffe



Auswertungsbeispiele

Allgemeine Informationen zur Nutzung des Web Servers

Zeitpunkt der Berichterstellung	31. Mai 1998
Zeitraum der Auswertung	1. Juli 1997 bis 20. Juli 1997
Anzahl Hits/Visits in diesem Zeitraum	77.165 Hits / 9.224 Visits
Gescheiterte Anfragen an den Server	0 (0,0%)
Trafficvolumen in diesem Zeitraum	198.624.728 Bytes
Mittlere Anzahl Hits/Visits pro Tag	3.948 Hits / 472 Visits
Mittlerer Traffic pro Tag	10.163.088 Bytes
Anzahl Dateien auf dem Web Server	488
Anzahl zugreifender Domains	3.136

Rangliste der 10 beliebtesten Seiten

Hauptdatei	Anzahl Hits	Gesamtanteil	Trafficvolumen
/infos/default.htm	2.638	15,4%	4.171.182 Bytes
/infos/deesse.htm	1.873	10,9%	14.579.727 Bytes
/infos/laties.htm	1.644	9,6%	18.763.540 Bytes
/infos/newig.htm	802	4,7%	2.602.592 Bytes
/surft/default.htm	685	4,0%	0 Bytes
/surft/germ.htm	649	3,8%	9.870.228 Bytes
/surft/internet.htm	560	3,3%	890.934 Bytes
/default.htm	426	2,5%	342.848 Bytes
/surft/calcs.htm	420	2,5%	545.806 Bytes
/surft/magazin.htm	361	2,1%	1.242.529 Bytes

3. Problematik von Proxy-Servern

- Proxy-Server speichern (häufig gefragte) Seiten zwischen
- Ziel: Verringerung des Datentransfervolumens bei Providern oder Firmen
- Proxy-Server sind zwischen den Nutzer und den „Original-Server“ geschaltet
- Variante: Local-Cache (temporary internet files)

Beispiel:

- Seiten mit Nachrichten ändern sich ständig, der Einsatz eines Proxy-Servers ist hierbei nicht sinnvoll
- Seiten mit bleibendem Inhalt müssen nicht bei jedem Zugriff vom eigentlichen Informations-Server abgerufen werden, da der Inhalt sich nicht verändert hat

→ www.Bibel.de Der Inhalt wird sich kaum verändern !

3. Problematik von Proxy-Servern

- Je nach Konfiguration überprüft der Proxy-Server Veränderungen
 - sofort (immer)
 - stündlich, täglich, monatlich, etc.
 - nie
- Zugriffe, die vom Proxy-Server bedient werden, werden nicht registriert
- Log-Files entsprechen nicht den tatsächlichen Anforderungen der Seite
 - Verfälschte Ergebnisse bei Leistungskennziffern

4. Neue Methoden der Leistungsmessung

Das **Fliege-Verfahren** umgeht die Proxy-Problematik

- Auf einer Web-Seite wird ein Abruf einer dynamischen Seite eingebaut
- Abruf / Ergebnis ist für den User nicht erkennbar (1x1-Pixel Grafik)
- Abrufe dynamischer Inhalte können nicht vom Proxy-Server bedient werden
- Eintrag im Log-File wird erzwungen
- Grundlage des IVW-Verfahrens zur ‚objektiven‘ Messung der Werbeträgerleistungen

4. Neue Methoden der Leistungsmessung

Durch den Einsatz von **Cookies** werden Rechner / User identifiziert

- Cookies sind Textelemente die auf dem Rechner des Users abgelegt werden
- Speicherung in speziellen Ordner oder Dateien
- Cookies werden vom Server gesetzt und abgefragt.
Erlauben -bei wiederholtem Zugriff auf eine Seite- die Identifikation des Rechner / User (nicht wer, aber wie oft)
- Durch Nutzung mehrerer User an einem Rechner wird ggf. nicht der User, sondern nur der Rechner identifiziert
- Das „Annehmen“ von Cookies kann durch Browsereinstellung unterbunden werden:
 - Cookies generell annehmen
 - Cookies generell ablehnen
 - Vor Cookies warnen

4. Neue Methoden der Leistungsmessung

Beispiel eines Cookies :

Set-Cookie:

```
B=7l6jsfotv4p92&b=2; expires=Thu, 15 Apr 2010 20:00:00 GMT;  
path=/; domain=.yahoo.com
```

- Dieser Cookie wird bei www.yahoo.de automatisch gesetzt
- Ein Cookie wird (theoretisch) beim ersten Seitenkontakt gesetzt
- Bei erneutem Seitenkontakt wird serverseitig nach einem vorhandenen Cookie gesucht
- Dieser gibt Aufschluss darüber, wann diese Seite das letzte mal besucht wurde, evtl. auch Name des Users, etc.
- Somit lässt sich ein Nutzerprofil erstellen, das Nutzerverhalten kann aufgezeichnet werden (Tracking)
- Cookies erlauben das Wiedererkennen von Usern und ermöglichen beispielsweise eine personalisierte Bannerwerbung (Targeting)

4. Neue Methoden der Leistungsmessung

- Web-Inhalte sind oft nur verfügbar, wenn das Setzen von Cookies erlaubt wird
- Trend: Ein „angenehmes“ Surfen ist nur bei „Cookies generell annehmen“ möglich
 - einige Nutzer löschen regelmäßig Cookies (manuell/automatisch)
 - Cookies besitzen ein schlechtes „Image“ (Angst vor Viren, Spionage, etc.)
- Voreinstellungen „vertuschen“ die „Cookie-Problematik“
- Zentrale „Datensammler“ (z.B. Doubleclick“) ermöglichen Site-übergreifende Analysen

4. Neue Methoden der Leistungsmessung

Erweiterung der „Fliege“: **Web-Bug** oder Clear GIFs:

- Beim Abruf der dynamischen Seiten werden zusätzliche Informationen an den Ziel-Server übermittelt. Die Daten werden durch Java-Script oder Activ-X Controls vom Rechner des User ermittelt
- Im Gegensatz zu Cookies bemerkt der User das Ausspähen nicht
- Web-Bugs können nur im Source-Code der Website entdeckt werden

4. Neue Methoden der Leistungsmessung

Beispiel eines Web-Bug:

```

```

```
<IMG WIDTH=1 HEIGHT=1 border=0  
SRC="http://media.preferences.com/ping  
?ML_SD=IntuitTE_Intuit_1x1_RunOfSite_Any  
&db_afcr=4B31-C2FB-0E2C  
&event=reghome  
&group=register  
&time=1999.10.27.20.5 6.37">
```

- Dieser Web-Bug findet sich auf der Startseite von www.quicken.com
- Das Pixel hat die Größe 1x1 und ist unsichtbar (transparent)
- Daten werden an DoubleClick und MatchLogic (preferences.com) weitergeleitet

4. Neue Methoden der Leistungsmessung

- Durch den Einsatz von Web-Bugs können folgende Informationen an den Server übermittelt werden:
 - Die IP-Adresse des Computers der den Bug aufruft
 - Die URL der besuchten Web-site
 - Die URL des Web-Bug Gif's
 - Den Zeitpunkt, an dem der Bug angeschaut wurde
 - Browsertyp
 - Die letzte Seite, von der aus der Aufruf erfolgte
 - Welche einzelnen Seiten innerhalb einer Web-Site besucht wurden und wie lange
- Daneben sind Web-Bugs in der Lage Cookies zu installieren und können u.U. auch Surfbewegungen aus dem Lokal-Cache auslesen

4. Neue Methoden der Leistungsmessung

- Durch Einsatz dieser Technik ist das Aufzeichnen eines Clickstreams möglich
- Im Gegensatz zu reinen Log-Files wird also nicht nur ein einzelner Aufruf, sondern eine ganze User-Session protokolliert
- Durch den kombinierten Einsatz von Web-Bugs und Cookies ist das Tracking von User möglich, also das Aufzeichnen und Wiedererkennen des Surfverhaltens und das Zuordnen zu einzelnen Usern
- Der Einsatz von Web-Bugs ist Datenschutzrechtlich höchst problematisch
 - Das Einverständnis des User ist ggf. nicht gegeben
 - Der User weiß nicht, welche Informationen weitergegeben werden
 - und an wen diese Informationen weitergegeben werden
- Viele Online-Werbeagenturen tauschen zudem Daten untereinander aus, bzw. gestalten ihre Cookies für andere Agenturen Transparent (nutzbar)

4. Neue Methoden der Leistungsmessung

- Eine einfachere und vollkommen legale Möglichkeit bietet das Erfragen von Nutzerdaten auf freiwilliger Basis: Personalisierte Dienste
- Web-Dienste sind teilweise nur verfügbar, wenn der User sich identifiziert
- Login bei Portalen / Suchmaschinen:
 - Maildienst bei My Yahoo (Yahoo-Portal)
 - ...
- Log-Files können mit User-Informationen kombiniert werden

4. Neue Methoden der Leistungsmessung

- Problematik der Richtigkeit der Angaben
 - Unerheblich für Daten in Bezug auf Seitenabrufe, etc.
 - Erheblich für Auswertungen mit soziodemographischem Charakter (Zielgruppenbildung)
 - Oftmals wird nur der Name verändert, das „Ausdenken“ falscher Geburtsdaten oder Berufe ist umständlich
- „Echte“ Nutzerdaten bei Internet-Provider (AOL, T-Online) oder Diensten, die eine Identifizierung voraussetzen:
 - Banken (Identverfahren Deutsche Post)
 - Web.de (Freischaltkennwort per Briefpost)

Abwägung: Schnelligkeit / Kosten gegen 100% Exaktheit

5. Werbeträgerforschung durch Institute

- Möglichkeit zur Erhebung von Internet-Daten mittels Online-Panels (z.B. PC-Meter der GfK Nürnberg)
 - Aufzeichnungsgeräte werden im Haushalt an den Rechner angeschlossen
 - Legale Datenerhebung
 - Fraglich in Bezug auf Validität:
 - Nutzungsverhalten gleich eines „Anonymen Users“? (xxx-Angebote)
 - Nur Aussagekräftig bei „großen“ Anbietern
- Online-Befragungen
 - Ohne Kopplung an Log-File Analyse (Generelle Befragung durch Mafo-Unternehmung)
 - Mit Kopplung an Log-File Analyse (Online-Befragung einzelner Seitenanbieter zur Erhebung soziodemographischer Daten)
- IVW: Leistungskennziffern von Werbeträgern