

Analyse des Nutzerverhaltens zum Videoserverprojekt der Friedrich-Schiller-Universität Jena

Norbert Kleekamp
Friedrich-Schiller-Universität
Jena, Multimediazentrum
Ernst-Abbe-Platz 4
07743 Jena
ank@uni-jena.de

Tino Tschiesche
Friedrich-Schiller-Universität
Jena, Multimediazentrum
Ernst-Abbe-Platz 4
07743 Jena
Tino.Tschiesche@uni-jena.de

Zusammenfassung

Seit dem Jahr 2000 werden an der Friedrich-Schiller-Universität Jena Streaming Technologien genutzt und seit 2001 werden Vorlesungen aufgezeichnet. Gegenwärtig sind ca. 4000 Videos weltweit abrufbar, ein großer Teil davon bereits mit synchron angebundene Präsentationen. Für die Beteiligten an dem Projekt ist wichtig zu untersuchen, wie das Projekt genutzt wird. Da diese Fragestellungen nicht mit der herkömmlichen Webmining-Software beantwortet werden konnten, wurde eine individuelle Lösung mit dem SAS-System erarbeitet. Um die Arbeit damit auch Mitarbeitern zu ermöglichen, die mit dem SAS-System nicht vertraut sind, wurde ein Projekt im SAS Enterprise Guide 3.0 erstellt.

Schlüsselwörter: Videoserverprojekt, Digitale Bibliothek Thüringen, Datenbank, SAS Enterprise Guide, Weblogmining, Webmining

1 Projektschwerpunkte Videoprojekt

- Das Projekt begann mit der Aufnahme einzelner Vorlesungen im Jahr 2001 als Pilotprojekt.
- Die Verbindung des Streaming Servers mit der Digitalen Bibliothek Thüringen ist Ergebnis eines DFG Verbundprojektes mit der Universität Essen (2001-2003).
- Weitere Verbesserungen durch Grafikanbindung (HWP Projekt 2003)
- Gegenstand dieses Projektes ist es, das Angebot von on demand Veranstaltungen wesentlich zu erweitern und die Nutzung vorhandener Aufzeichnungen durch Grafikanbindung und Nachbearbeitung zu verbessern.
- Zugriffszahlen seit Januar 2006: ca. 215.000 Zugriffe, 43.200 h Abrufzeit, Datentransfer 3,6 TB, 4000 Videos im Bestand, davon 2000 in DBT
- 77% der Zugriffe erfolgen von außerhalb des Netzes der FSU Jena.

2 Realisierung mit dem SAS Enterprise Guide

Für die Dozenten und Administratoren ist es wichtig, zu wissen, wie und von wem der Dienst genutzt wird. Mit Hilfe von Statistiken, Tabellen und Grafiken soll diese Fragestellung fundiert beantwortet werden.

Da das SAS -System mit dem Enterprise Guide zur Verfügung stand und es klar war, dass individuelle Lösungen mit den üblichen Tools nicht realisierbar sind, wurde entschieden, ein Projekt mit dem SAS-System zu erarbeiten.

Jeder Zugriff auf den Videosever erzeugt in einer log-Datei einen Eintrag. Die Log-Dateien der FSU Jena enthalten gegenwärtig ca. 4 Mio. Datensätze/Jahr, von denen 10% relevant für eine Zugriffsstatistik sind. Hier ein Beispiel:

```
141.35.179.219 - - [19/Feb/2007:18:39:45 +0100] "GET DBT/61/26/061124-153625-611-7054_10174.rv?cloakport=80,554 RTSP/1.0" 200 164221136
[MacOSX_10.4_10.0.0.0_play32_RN01_EN_686_No-FPU] [7582f161-bc21-11db-e8e0-7116d353a5cb] [UNKNOWN] 264923252 5920 8241 1208 12964 5188
```

Aus diesen Daten werden IP-Adresse, Dateiname, Startzeit, gesendete Bytes und Abrufdauer ausgelesen. Weitere Metadaten bezüglich Videonamen, Autoren, Kategorien werden aus den Datenbanken der Digitalen Bibliothek Thüringen (www.db-thueringen.de) extrahiert und mit den Daten aus der Log-Datei verknüpft.

Diese Aufgaben werden im Projektcode „Dateneingabe“ gelöst, es entsteht eine SAS-Tabelle mit allen relevanten Informationen für jeden Zugriff. Es schließt sich eine Abfrage an, in der der Zeitbereich eingestellt und eine Videoauswahl getroffen werden kann.

Ein weiterer Projektcode „Auswertungen“ erzeugt SAS-Tabellen für die Anzahl der Zugriffe, gesendeten Bytes und Abrufzeiten in verschiedenen Zeitintervallen (je Stunde, je Tag und je Woche). Jetzt können mit den Anwendungsroutinen des Enterprise Guide eigene Grafiken und Tabellen erzeugt werden, z.B. Entwicklung der Zugriffe, der gesendeten Bytes und Abrufzeiten insgesamt und getrennt nach internen bzw. externen Nutzern.

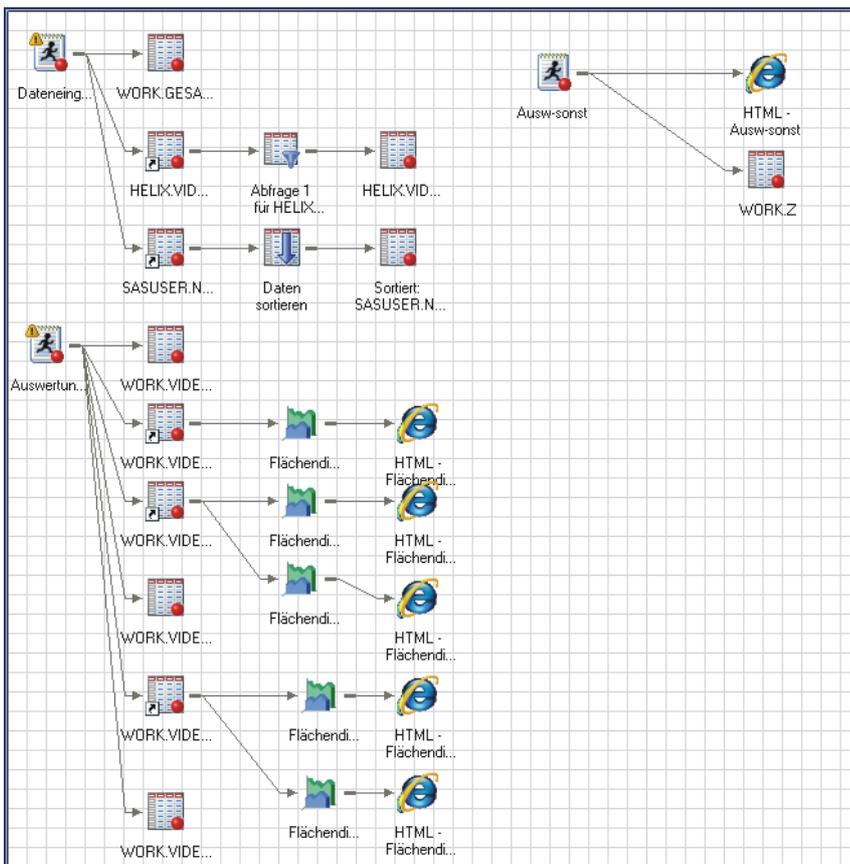


Abb. 1 Prozessflussfenster aus dem EG

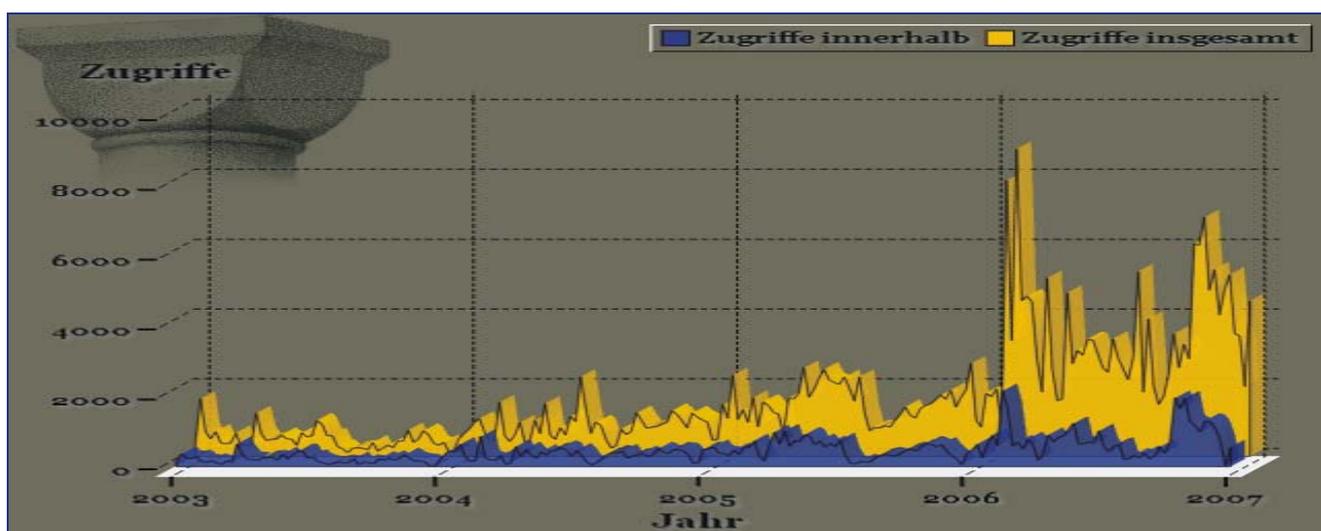


Abb. 2 Zeitlicher Verlauf der Zugriffe über mehrere Jahre

Im Projektcode „Ausw-sonst“ werden weitere Tabellen erstellt. Diese sind jeweils nach Namen, Häufigkeiten (Hitliste) bzw. Nutzern oder Datum sortiert.

- Anzahl Zugriffe je Titel
- Abrufzeiten je Titel
- Abgerufene Datenmenge je Titel
- Nutzerstatistik (Zugriffe je Location)
- Letzter Abruf des Titels

	Location	Anzahl
1	FSU Jena	47846
2	Telekom dialin	21051
3	Arcor dialin	16170
4	84.184	9976
5	84.59	6980
6	82.119	4008
7	Freenet dialin	3782
8	84.151	3617
9	Japan	3064
10	AOL dialin	3047
11	Tiscali dialin	2958
12	84.185	2594
13	70.83	2446
14	84.177	2380
15	TU Ilmenau	2254
16	84.182	1950
17	FH Erfurt	1905
18	83.129	1697
19	84.191	1671
20	90.186	1479
21	84.57	1326
22	FH Wernigerode	1280
23	84.179	1196
24	84.174	1192
25	TU Dresden	978

Abb. 3 Nutzerstatistik

Beob.	Titel	Summe send_time
1	Smabs2004_workshop_muthen_22_1.rm	440:42:45
2	4. Reziprozität- Kulturanthropologie	359:44:56
3	Semantic Web (01) - Die Grenzen des	341:55:17
4	3. Kulturökologie- Kulturanthropologie	324:51:43
5	Interkulturelle Wirtschaftskommunikation	311:54:35
6	Interkulturelle Wirtschaftskommunikation	295:24:52
7	01 Methoden der Werbeanalyse-	294:03:43
8	5. Religion- Kulturanthropologie	288:59:30
9	Theorien psychometrischer Tests	288:25:54
10	Methoden der Evaluationsforschung	281:46:41
11	Smabs2004_workshop_vermunt.rm	275:10:24
12	8. Medien und Tradierung-	260:20:45
13	7. Identität- Kulturanthropologie	260:00:10
14	Interkulturelle Wirtschaftskommunikation	256:09:33
15	Interkulturelle Wirtschaftskommunikation	255:02:04
16	Hinweis.rt	253:40:07
17	Theorien psychometrischer Tests	252:52:09
18	Interkulturelle Wirtschaftskommunikation	249:59:28
19	2. Anthropologie- Kulturanthropologie	248:22:16
20	Methoden der Evaluationsforschung	243:42:13
21	10. Münch I: Handlungsfelder-	237:34:37

Abb. 4 Hitliste nach Abrufzeit