Monitoring User Pattern in School Information System Using Logfile Analysis

Arne Hendrik Schulz, Andreas Breiter
Overview

• Research Questions
• Project Design
• School Information Systems
• Logfile Analysis/Web Mining
• Data
• Results
• Conclusions
Research Questions

• How do digital media change communication among teachers within schools?
  • Focuses not on communication among teachers and pupils!
  • School research in recent years did not focus on this point
• What does effect higher exchange rates among teachers?
• Part of the SPP Priority Program „Mediatized Worlds“
Research Agenda

- Logfile analysis
- 2 schools
- Group discussions
Mediatization
School Information Systems (SIS)

• (mostly) webbased systems to support teachers work

• Include
  • Resource planning
  • Groupware (teachers, pupils, parents)
  • Decision support
  • Pupil management (classes, marks)
  • Learn Management Systems (LMS)
Logfile analysis

• Technical origin
• Logging at server level or application level
• Example logfile (application level)
  • Home Mozilla/5.0/Ubuntu 2012-01-19 06:42:37 /url/site.php?mod=announcement&fct=detail&iid=3041791 GET 1242348
  • Home IE/8.0/Windows7 2012-01-19 06:42:44 /url/site.php?mod=material&fct=index GET 2164158
  • School Mozilla/5.0/MacOSX 2012-01-19 08:16:46 /url/site.php?mod=home&fct=index GET 2677255
Web Mining

• Subgroup of Data Mining
• Methods for dealing with web data
• 3 main categories
  • Web Structure Mining
  • Web Content Mining
  • Web Usage Mining
Web Mining aims

• Web Structure Mining
  • Hyperlink-ranking/sitemaps
  • Web search

• Web Content Mining
  • Website clustering/classification
  • Information retrieval/Web knowledge

• Web Usage Mining
  • User clustering/classification
  • Personalization
  • E-Business (Commerce, Marketing, …)
Data (I)

- March 2011 until March 2012
  - Interruption between Mid of April and Mid of May; middle of September
- 120,000 hits
- ~60,000 hits after data cleaning and completion
- 138 different user
- 4,451 visits
Data (II)

- 650 materials
  - One material may contain more than one file
- 660 dates
- 275 announcements
- > 3,000 files
SIS access
Access to different modules

- Most accesses on Sunday and Monday
- Materials, dates and announcements used most often
- Frequent use of search functions
3 main user groups
Visual SIS-Structure

- As social network
  - Items as nodes
  - Hits as size
  - Links as edges
  - Item type as vertex color
- Force-driven algorithm (Fruchterman-Reingold)
  - Closer linked items grouped
  - Frequently used links between items group
- Web Structure Mining and Web Usage Mining
Visual SIS-Structure

- Miscellaneous
- Exams
- Reports
- Conferences
- Topics
  - Spanish
  - English
Materials per topic
Visual SIS-Structure

- Social science, Mathematics, Business studies and Biology are close together
- Materials are linked to multiple topics
Materials per topic during research
Who uploads to thick topics

- Mostly contribution to one topic
- One heavy user uploads to almost any topic
  - Assistance?
- Two other heavy user to not contribute much topic related materials
- English and Spanish have larger „contribution groups“
- Spanish uploads mostly by one teacher
Conclusions

• SIS mainly used for announcements, dates and materials
• Closely connected SIS
• 3 groups of users
• Contribution of materials
  • Topics: English, Spanish, Social Science
  • Characterized by individualistic vs. Collective
  • Topic-specific
Remarks

• Communication by other systems (E-Mail, USB, Dropbox etc.) not covered
• Still a lot „offline“ communication
  • Mailboxes/notes
  • Printed materials
• Users‘ aims
  • Logfiles do not have these informations
• Number of computers/state of network in school low
Thank you very much for your attention!