Web Analytics with Formal Concept Analysis

Sanda Dragoș and Christian Săcărea
Formal Concept Analysis

0 Data Analysis, Knowledge Management
0 Basic data structure - formal context (G,M,I): G set of objects, M set of attributes, I ⊆ G x M, gIm “object g has attribute m”.
0 Knowledge extracted with the help of a Galois connection between G and M, known as derivation operators.
0 A ⊆ G, A′ := {m ∈ M | gIm for all g ∈ A}.
0 B ⊆ M, B′ := {g ∈ G | gIm for all m ∈ B}. 

30.05.2012 
Zilele Academice Clujene, 2012
Conceptual hierarchies

0 Formal Concept: \((A,B), A'=B, B'=A\)

0 Subconcept-superconcept relation: \((A,B) \leq (C,D)\) if and only if \(A \subseteq C\) (which is equivalent to \(D \subseteq B\)).

0 Complete lattice: conceptual hierarchy.
Many-valued contexts \((G,M,W,I), (g,m,w) \in I\) “object \(g\) has attribute \(m\) with values \(w\)”. 

Typical database situation

Needs interpretation via a process called conceptual scaling
TOSCANA

- Knowledge Management suite
- Enables knowledge navigation, scenario building, visualization, diagram connection
- Connect to a database
- Use Elba to explain the semantics of tables in form of conceptual hierarchies
- Start Toscana to browse knowledge
Formal Concept Analysis on web usage data

- Web usage data:
  - from an e-learning portal called PULSE
  - over the last two months (i.e. April and May)

- Analysis:
  - Who is using PULSE? In what proportion?
  - Teaching subjects
  - Accessed information
  - ... and more
Who is using PULSE?

- 73.95% students = 8334 accesses
- 3.85% teacher = 434 accesses
- 22.2% pre-login = 2502 accesses
Years of study

- 68.36% current = 7704
- 3.39% past = 382
- 28.25% others = 3184
Teaching subjects for students

- Previous semester:
  - AC = 52
  - PW = 21
  - WP = 7

- Current Semester: 71.93%
  - SO1 = 8107
Teaching subjects for teacher

- Previous semester: 0
  - AC = 0
  - PW = 0
  - WP = 0

- Current Semester: 94.93%
  - SO1 = 412
PULSE for students

- 23.11% labs = 2605
- 17.37% lectures = 1958
- 2.22% logout = 250
- 0.69% news = 78
- 0.35% feedback = 40
- 0.2% faq = 22
PULSE for students
- details on labs -

- 53.05% assignments = 1382
- 42.03% lab = 1095
- 4.91% examples = 128
PULSE for students - details on lectures -

- 57.3% lecture material (Ro) = 1122
  - 27.07% examples = 350
- 12.97% test papers = 254
- 2.66% lecture material (En) = 52
PULSE for teachers

- 39.4% markAssignments
- 11.29% lectures
- 11.06% markAttendees
- 5.07% labs
- 4.38% addAssign
- 3% studentList
- 1.61% lecturePaper
- 0.92% addNews
Nested diagram: months & login

- Mai:
  - 3577 students
  - 195 teacher

- Aprilie:
  - 4757 students
  - 239 teacher
Conclusions

- Formal Concept Analysis
- a new approach into web analytics