

Association Rules Mining from Context-enriched Server Logs

Supervisor: Dušan Zeleník
Juraj Višňovský

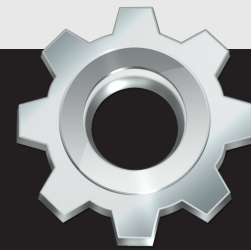
Adaptive Proxy Server

PeWe proxy server improves users' browsing experience by personalizing Web content.
» over 3 million access logs
» 77 unique users



Data Preprocessing

» Logs with no user identification elimination
» Non-human activities detection and removal
» Sessions identification



Goals

- Web users' habits identification
- Confirm context's influence on Web users' activities
- Compare association rules generation for clear logs and context-enriched logs
- Interesting knowledge discovery

Contexts

Any relevant information used to describe a state of a user. Activity of the Web user is influenced by many different contexts, such as:

- » Time
- » Location
- » Occupation
- » Weather
- » Web domain category



Prediction

Predictions of Web user activities are based on knowledge discovered from frequent patterns. Predictions are represented by a set of association rules.

Recommendation

Real-time recommendation for visiting Web domain or specific Web site. Based on past activities recorded by PeWe Proxy server.

Frequent Patterns

» FP-Growth algorithm is used to generate frequent patterns
» Large datasets are processed with Hadoop using Apache Mahout machine learning library