

Exploiting Taxonomies' Broadening for Enhanced Tagging: Damped Concept Clustering

Christoph Wieser (Salzburg Research)

François Bry (University of Munich)

Århus, 18.9.2008



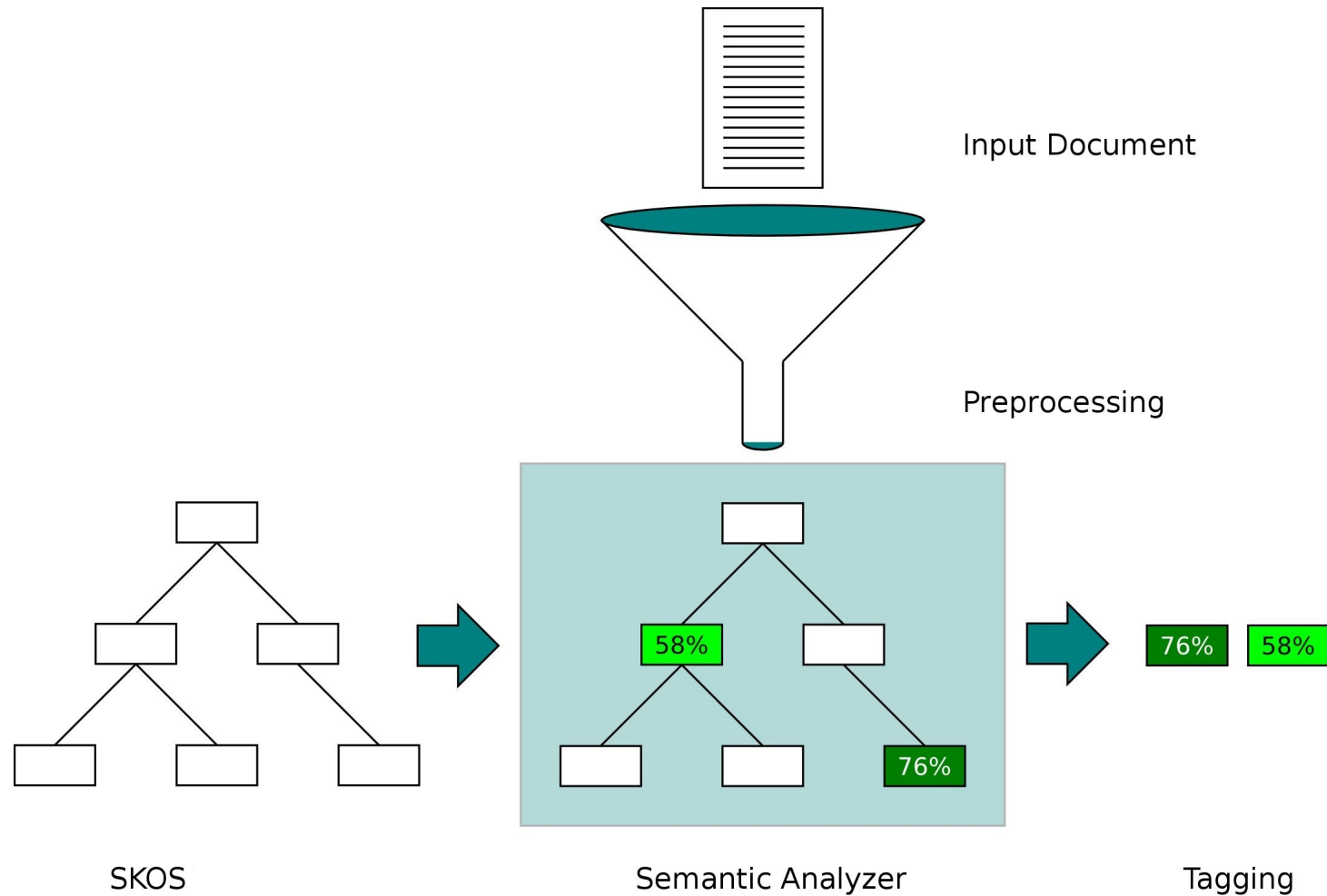
"Newspaper Editors don't tag anything"

- Project "Future Content Platforms": What does the newspaper of the future look like?
- E-Paper, Citizen Journalism, Tagged Articles

The screenshot shows a news website interface. At the top, there are navigation tabs: Nachrichten, Sport, Lifestyle, Video, Meinung, meinSalzburg, Motor, Immo, Karriere, Gastrojobs, Partnersuche. Below the tabs is a search bar with a red border and buttons for 'suchen' and 'löschen'. To the right of the search bar is a link for 'Personensuche in Internet Erweiterte Suche'. Below the search bar, there are three article snippets under the heading 'Artikel'. Each snippet includes a title, a short description, and a 'mehr...' link. The first article is titled 'Küchenfreunde' and mentions 'Schwab Küchen mit neuer Website'. The second is 'Platz in der Küche ist kostbar, verschenken Sie ihn nicht' and mentions 'Bayerhammerstraße 25'. The third is 'CarlaBruniTagundNacht' and mentions 'Carla Bruni-Sarkozy'. To the right of the article snippets is a map titled 'Map vollansicht' with buttons for 'Karte', 'Satellit', and 'Hybrid'. The map shows a geographical area with various cities and countries labeled, including London, Paris, Berlin, and Rome.

Preprocessing

- Take the nouns of a given text.
- Match each noun against each concept of a SKOS model.
 - A noun and a concept match, if they are equal else if they are synonyms.
- Count each match.



Where are the relevant tags?

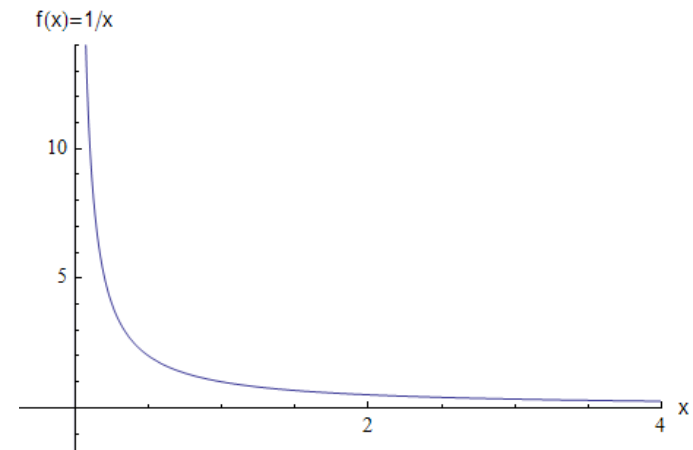
- Candidates: most often matched concepts
 - But we tend to use generalizations
- Let's try to detect the best generalizations and take them as proposal for tags.

Support of a Concept and Damping

- Support
 - Idea: A concept is a good generalization, if there are *narrower* concept matches.
- Damping
 - The farther a concept the lower its support.

Damped Concept Clustering

- Damping Factor
 - x is the Manhattan Distance between two concepts
- The support of a concept is the sum of the damped narrower concept frequency.
- Each concept represents a cluster.



Results

- Newspaper editors still don't tag.
- Extraction of
 - additional tags
 - semantic tags (generalization)
- Simple implementation
- Prototypical implementation and tests with users

Salzburg NewMediaLab

■ Website

- www.newmedialab.at

- pp.newmedialab.at

■ Contact

- Christoph Wieser
Salzburg NewMediaLab
c/o Salzburg Research
Jakob Haringer Straße 5/III
T. +43.662.2288-426

christoph.wieser@salzburgresearch.at

