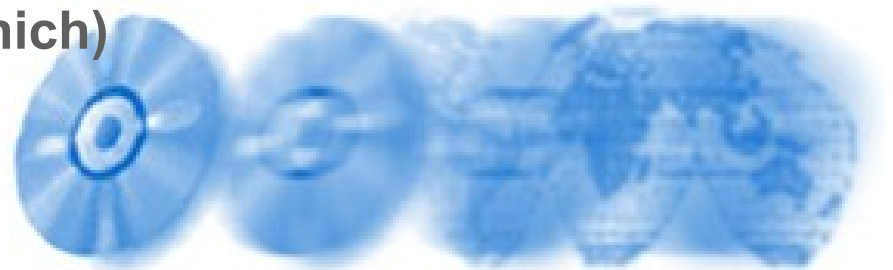


Exploiting Taxonomies' Broadening for Enhanced Tagging: Damped Concept Clustering

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 for "Nachrichten", "Sport", "Lifestyle", "Video", "Meinung", and "meinSalzburg". To the right, there are more tabs: "Motor", "Immo", "Karriere", "Gastrojobs", and "Partnersuche". Below the navigation is a search bar with "suchen" and "löschen" buttons, and a link for "Personensuche in Internet Erweiterte Suche".

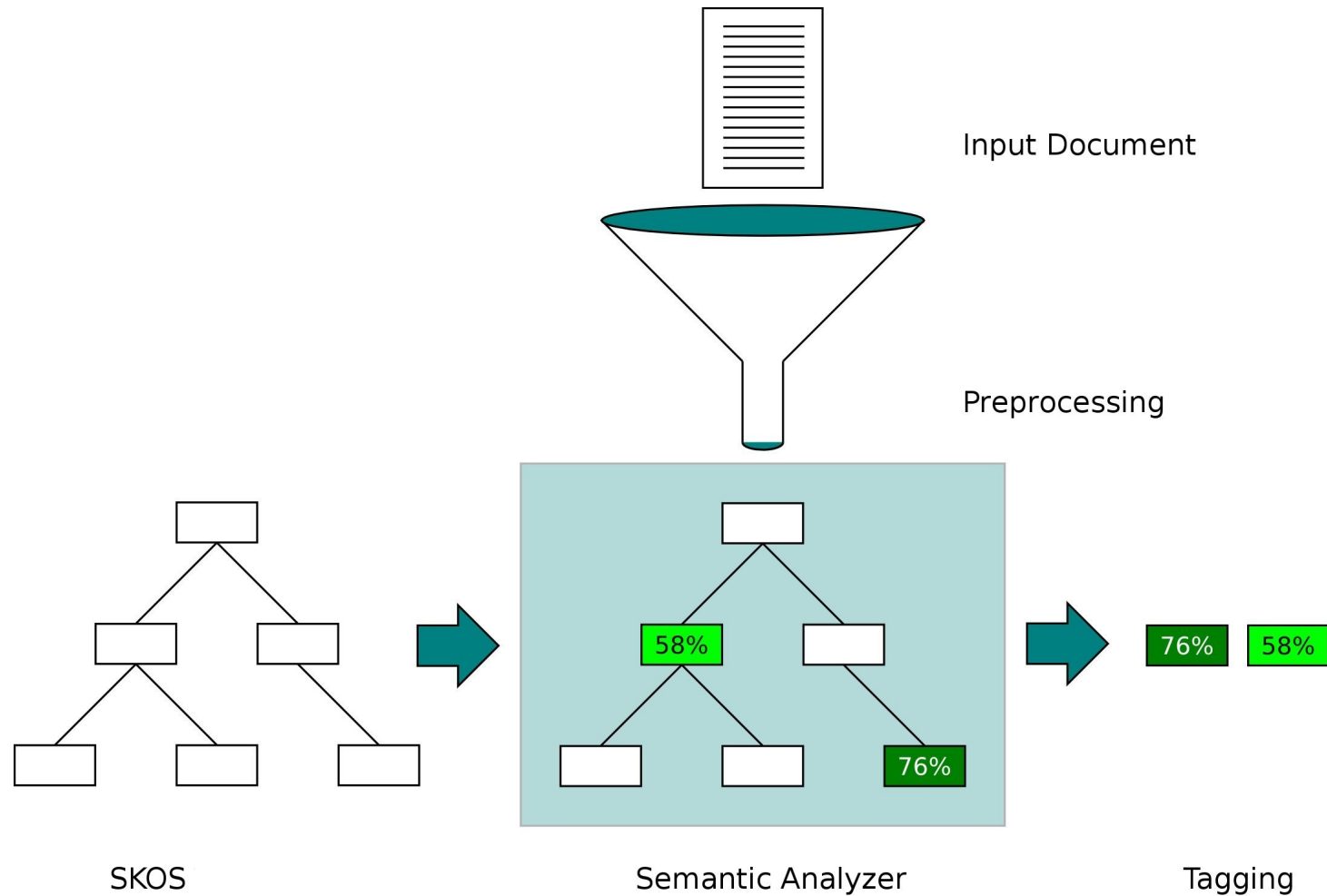
The main content area is titled "Artikel" and lists three articles:

- Küchenfreunde**
Schwab Küchen mit neuer Website: Neue Medien bestimmen zusehends unser alltägliches Leben, daher haben wir auch unsere
mehr...
Beilagen / 18.09.2008 18.09.2008 / [Print](#)
- Platz in der Küche ist kostbar, verschenken Sie ihn nicht**
„Wenn Ihre Küche so aussieht, wie die Ihrer Nachbarn, liegt es nicht an uns“ – meint das Team von Küchen Schwab in der Bayerhammerstraße 25. mehr...
Beilagen / 18.09.2008 18.09.2008 / [Print](#)
- CarlaBruniTagundNacht**
LONDON (SN, AFP). Die französische Präsidentengattin Carla Bruni-Sarkozy ist mit einem Song ihrer jüngsten CD live im BBC-Fernsehen aufgetreten und hat mit Moderator Jools Holland über Liebe und Sucht geplaudert. Ihr Lied „Tu es ma came“ (etwa: „Du bist mein Speed“) handle von „Abhängigkeit in der Liebe“, führte die Sängerin am Dienstagabend auf Englisch aus.
mehr...
TV / 18.09.2008 18.09.2008 / [Print](#)

On the right side, there is a "Map" section titled "vollansicht" with buttons for "Karte", "Satellit", and "Hybrid". The map shows Europe with various cities marked, 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

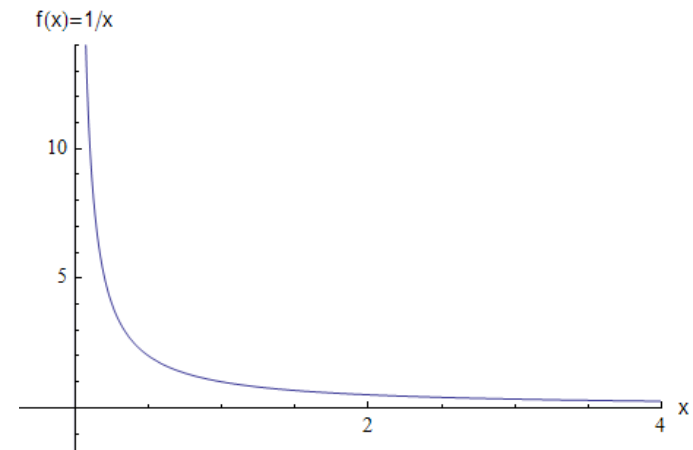
- Idea: A concept is a good generalization, if there are *narrower* concept matches.

- 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

