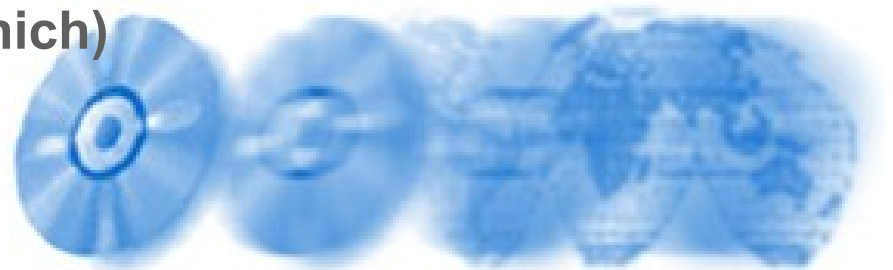# Exploiting Taxonomies' Broadening for Enhanced Tagging: Damped Concept Clustering

Christoph Wieser (Salzburg Research)

François Bry (University of Munich)

Århus, 18.9.2008

Industrielle
Kompetenzzentren und Netzwerke

EIN PROGRAMM DES BMWA

Christoph Wieser, Salzburg NewMediaLab
christoph.wieser@salzburgresearch.at

"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

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.

Input Document

Preprocessing

SKOS

Semantic Analyzer

Tagging

58%

76%

76% 58%
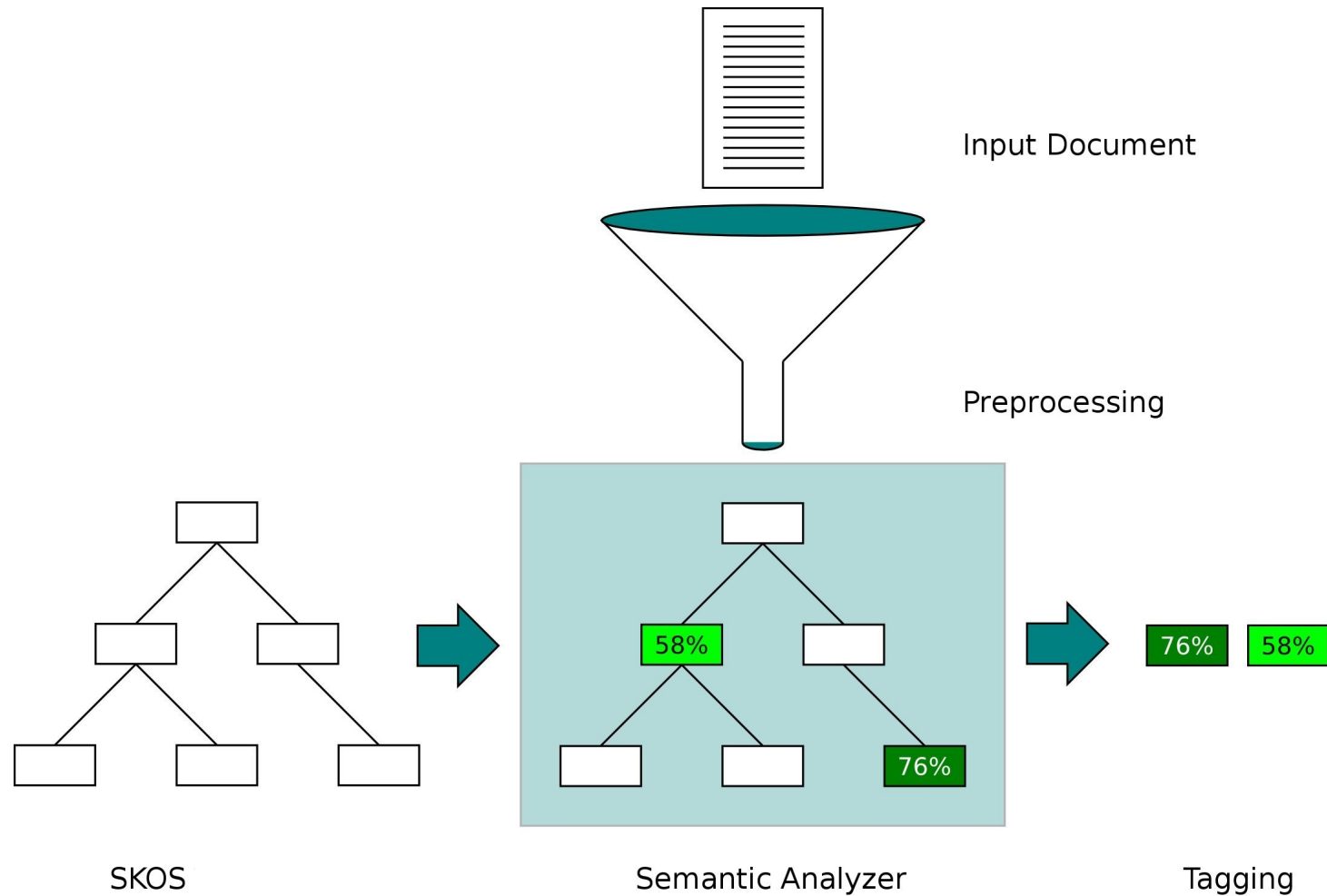
Where are the relevant tags?

- 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.

SNML
salzburg
newmedialab

# Support of a Concept and Damping

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

- **Damping**
  - The farer 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.

f(x)=1/x
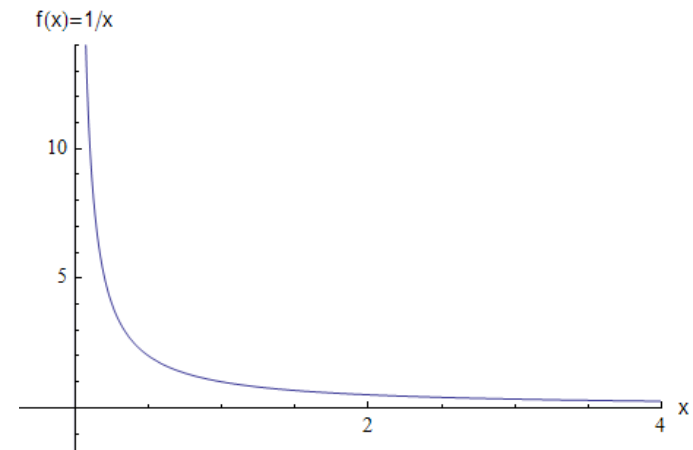
Results

- Newspaper editors still don't tag.
- Extraction of
  - additional tags
  - semantic tags (genaralization)
- 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