

# Text mining and visualization using VOSviewer

**Nees Jan van Eck** and Ludo Waltman  
Centre for Science and Technology Studies, Leiden University

1st Global TechMining Conference, Atlanta, USA  
September 14, 2011



Universiteit Leiden

# Outline

- Introduction to VOSviewer
- Text mining using VOSviewer
- Demo
- Applications



# VOSviewer

1. Tool for creating maps
  - Uses the VOS mapping and clustering techniques to create maps based on bibliographic data
  
2. Viewer for visualizing and exploring maps
  - Provides different views of a map
  - Provides zoom, scroll, and search functionality
  - Uses a special algorithm to prevent overlapping labels





# Text mining using VOSviewer

- Identification of terms in a text corpus
- Filtering of terms
  - Number of occurrences
  - Term relevance
  - Manual selection
- Construction of a term map
  - Mapping
  - Clustering



# Identification of terms

- Part-of-speech tagging
  - Identification of verbs, nouns, adjectives, etc.
  - Based on Apache OpenNLP
- Identification of noun phrases
  - Sequences of nouns and adjectives ending with a noun
  - For instance *paper, visualization, interesting result, text mining*
  - But not *degrees of freedom, highly cited publication*
- Transformation of plural into singular



# Filtering of terms based on relevance

- Relevant terms: *visualization, text mining, natural language processing*
- Non-relevant terms: *paper, interesting result, new method*
- Filtering of terms based on term co-occurrence network
  - For each term, the distribution of (second-order) co-occurrences is determined
  - The difference between this distribution and the overall distribution is calculated (using the Kullback-Leibler distance)
  - The larger the difference, the higher the relevance of a term

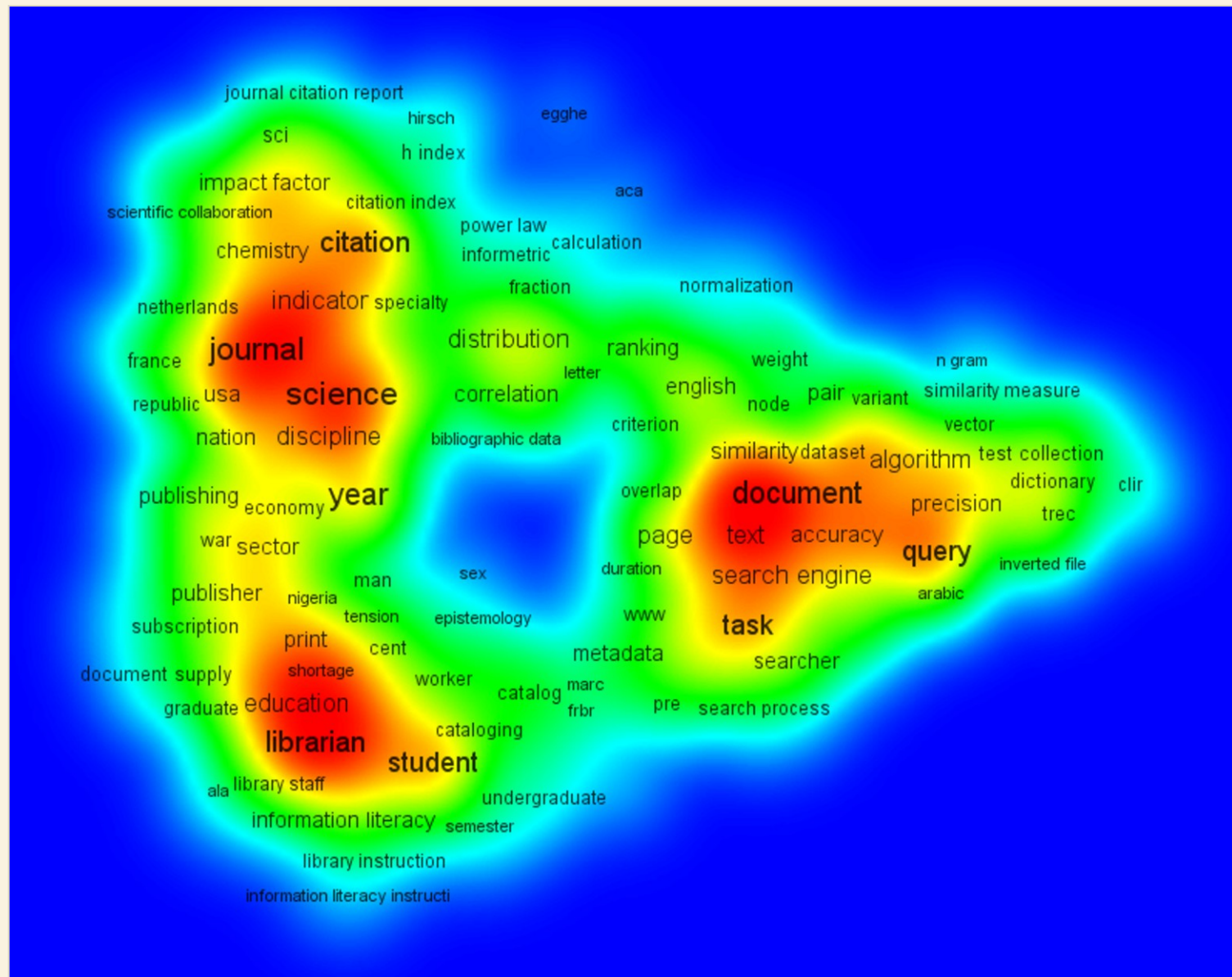


# Demo

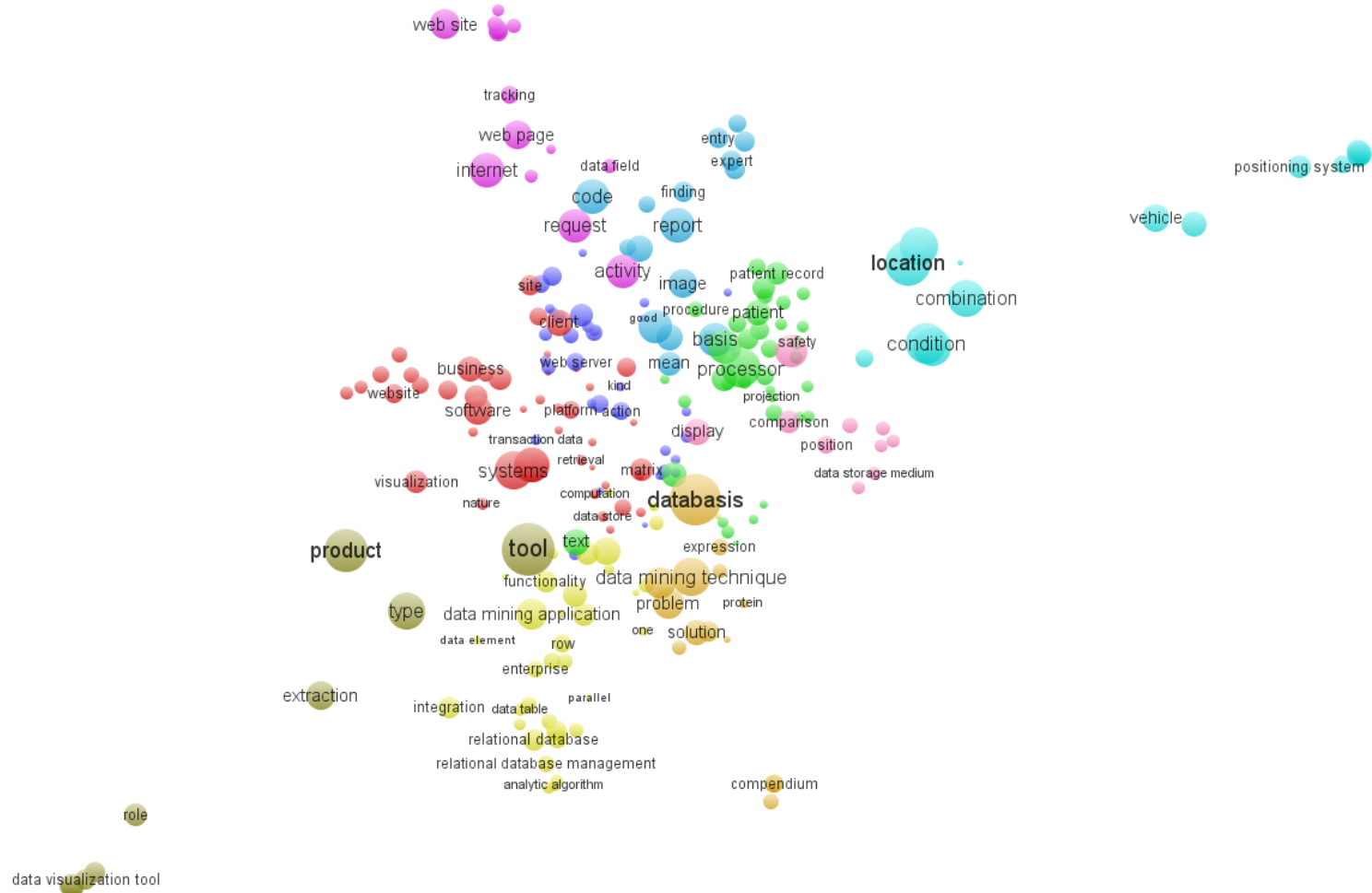
- Information science publications
  - Database: Web of Science
  - Titles and abstracts of 10,000 publications
- Data and text mining patents
  - Database: PATSTAT
  - Titles and abstracts of 1800 patents
- PhD thesis on science mapping
  - Full text analysis
  - Co-occurrences of terms counted at the level of paragraphs



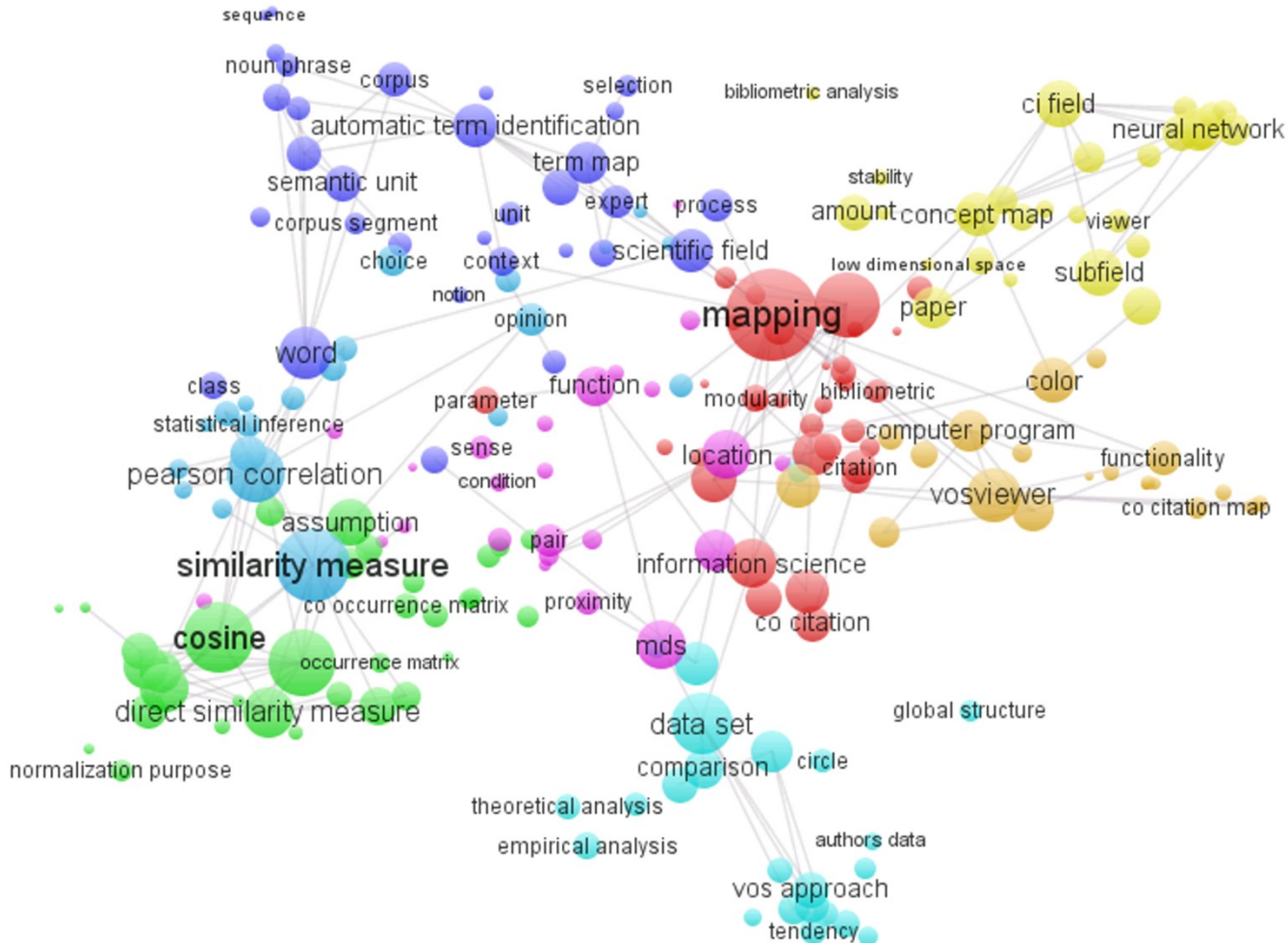
# Information science publications



# Data and text mining patents



# PhD thesis on science mapping



# Applications

- Areas of interest of Leiden University in information science research
- Identification of highly cited topics in the *Journal of the American Society for Information Science and Technology*







# VOSviewer website

[www.vosviewer.com](http://www.vosviewer.com)

- Download VOSviewer (free!)
- Examples of maps
- Publications



**Thank you for your attention!**

