



Google's Sense of Semantics

Kaser Armin, Kolar Gerald & Sachse Pierre

Google-Suche Auf gut Glück!

University of Innsbruck, Austria Department of Psychology
 Contact: Armin.Kaser@gmx.at



1. Introduction

Modern search engines display parallels to the functioning of the human brain (cf. Griffiths, Steyvers & Firl, 2007). Semantic distance may be inferred from the amount of hits which two terms obtain in a query. Thereby, semantic networks can be automatically constructed.

The present study aims to show:

- 1) which search engine may be the best for the construction of full semantic networks;
- 2) if web catalogues are better suited than search engines for the purpose of constructing semantic networks.

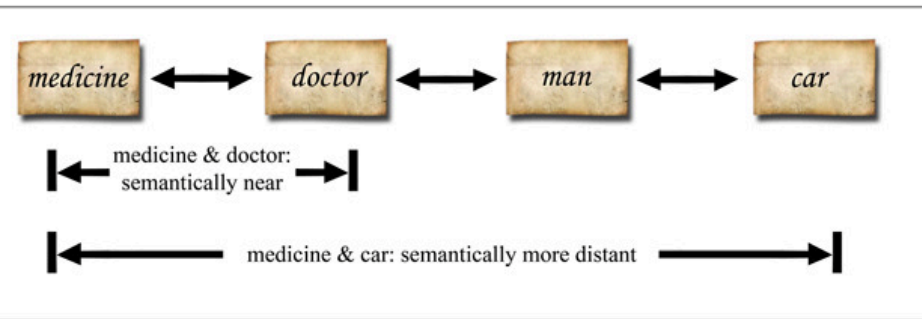
Whereas in search engines a robot autonomously includes new web-sites into the query, a web catalogue derives from the efforts of human beings compiling the information.

2. Methods

Starting point of the study was a natural (i.e. "human") data set amassed over several years. About 10,000 times two terms at a time have been selected from this data set in order to correlate their semantic distance with that of the search engines.

Phase 1: Semantic distance in the data set

Two terms being directly connected in the data set are semantically nearer than two terms being only connected via other terms in between.



Phase 2: Semantic distance in the search engines

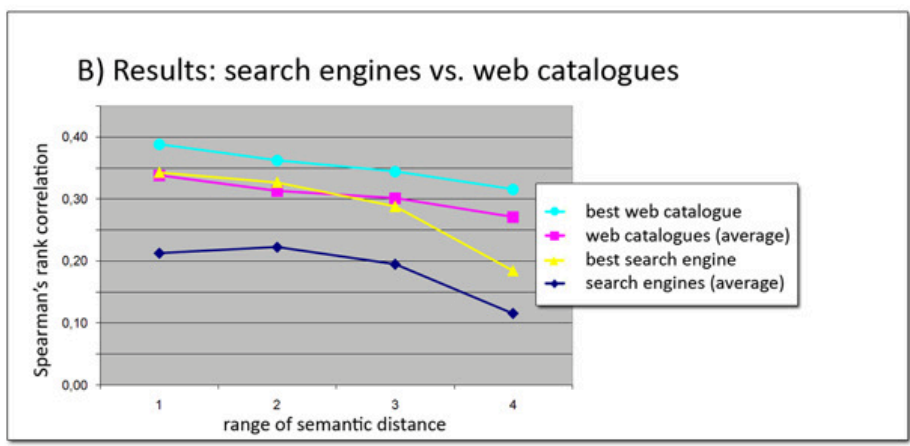
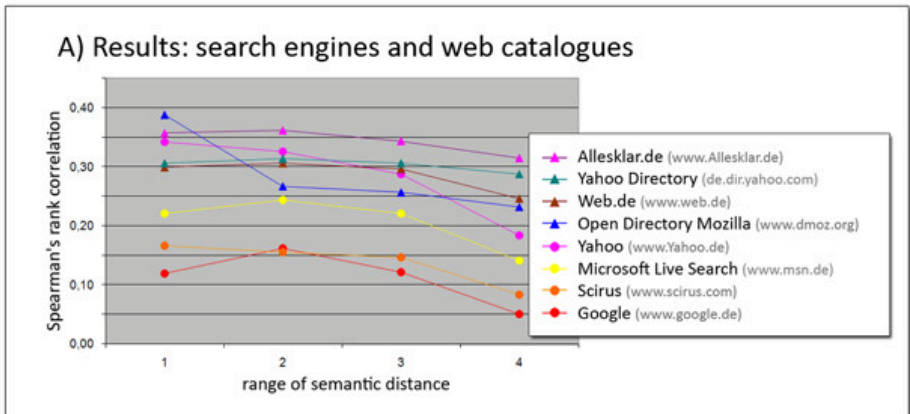
If two terms are entered into a search engine, they together obtain a certain amount of hits. Should those two terms be semantically near, they obtain more hits than terms that are more distant.

Phase 3: Correlation "semantic distance in data set – semantic distance in search engines"

The correlation gives information on how well search engines may be suitable for automatically constructing semantic networks.

3. Results

All results are significant.



4. Discussion

1) Results by categories

The examined search engines differ in their suitability of constructing semantic networks. The best results are:



category web catalogues:
 Allesklar.de (www.allesklar.de)



category search engines:
 Yahoo.de (www.yahoo.de)

2) Search engines vs. web catalogues

Web catalogues conceived by humans clearly obtain better results than search engines which have their indices automatically compiled by bots. However, search engines may have an advantage due to their substantially larger indices if the required terms are quite rare and therefore not included in web catalogues.