

Philipp Singer¹, Thomas Niebler², Markus Strohmaier¹, Andreas Hotho²

¹ Knowledge Technologies Institute, Graz University of Technology, Austria, ² Data Mining und Information Retrieval Group, University of Würzburg, Germany
E-mail: philipp.singer@tugraz.at, thomas.niebler@uni-wuerzburg.de, markus.strohmaier@tugraz.at, hotho@informatik.uni-wuerzburg.de

1. Introduction

Problem definition

Compute semantic relatedness between two Wikipedia concepts

Research Questions

1. Are navigational paths useful for calculating semantic relatedness?
2. Which kinds of navigational paths are particularly useful?

Contributions

We show that

1. human navigational paths provide a viable source for calculating semantic relatedness between concepts in information networks
2. specific subsets of paths can increase the quality of semantic relatedness calculations

Dataset "TheWikiGame"

Collection of 1.8 million click paths on Wikipedia

Navigation from a given start page to a given target page

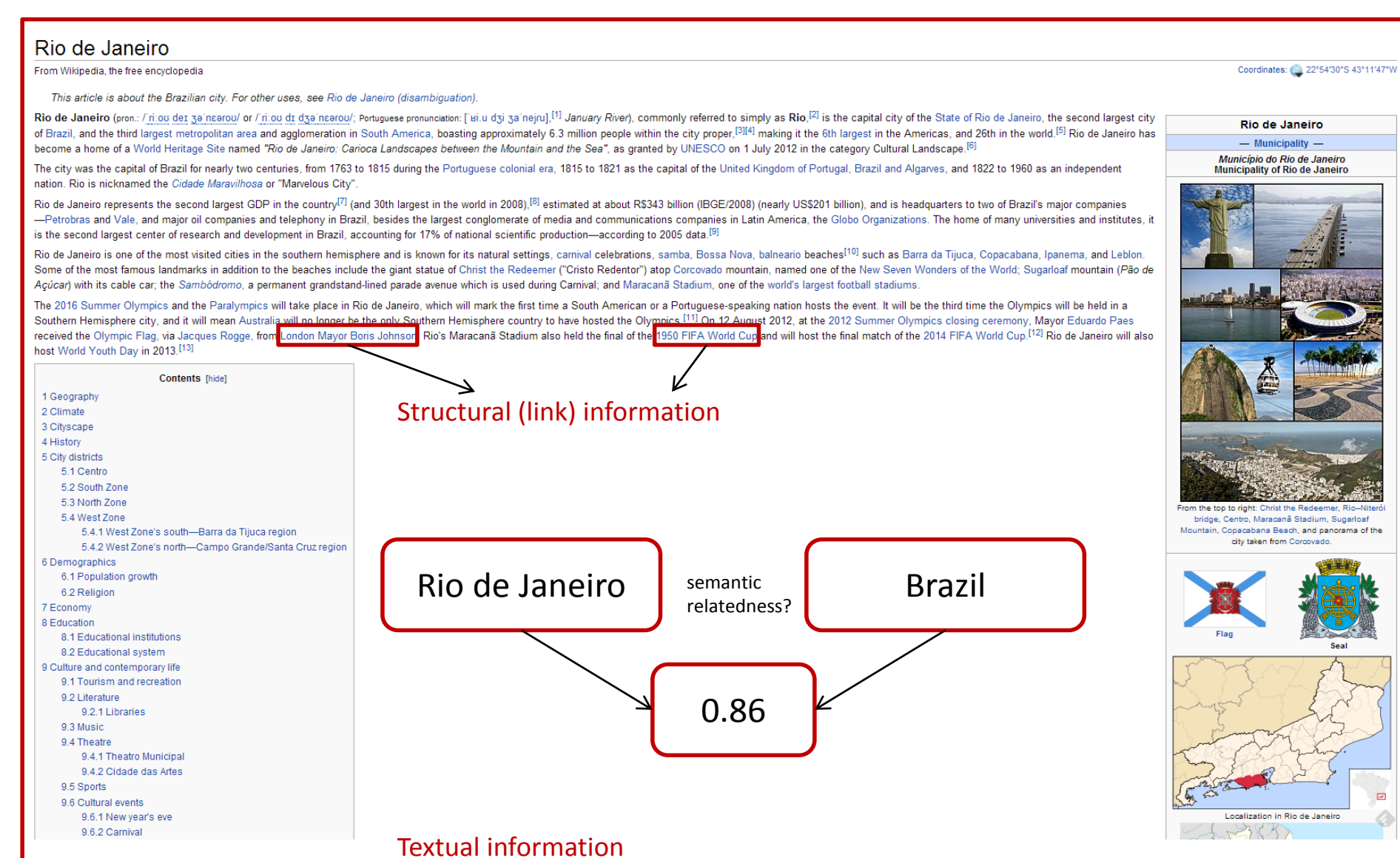


Figure 1: Examples of information used on Wikipedia for calculating semantic relatedness - in this work we focus on navigational paths through the link structure

2. Methodology

Calculating semantic relatedness on paths

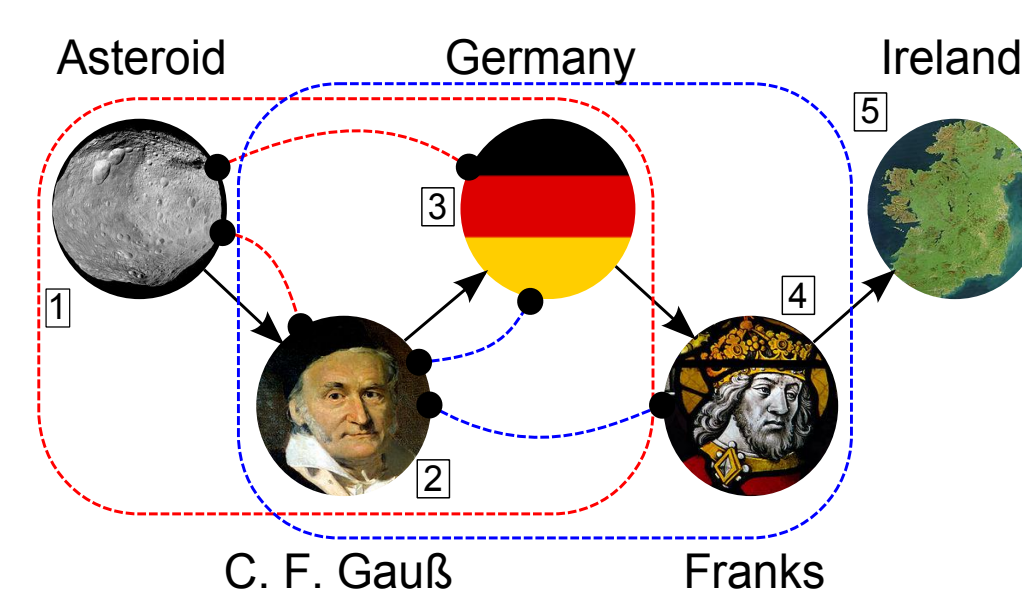
- First order co-occurrence information through sliding windows
- Cosine similarity between co-occurrence vectors

Evaluation

- Use WordSimilarity 353 data set [Finkelstein2002]
- 353 pairs of English words and names, mapped to Wikipedia concepts, similarity rating added
- Spearman Rank Correlation between WordSim353 and our results

Path Selection Experiments

- Select smaller path subsets according to certain path features
- Ten accumulating subsets towards the total number of visited nodes
- Calculate Spearman Rank Correlation for each path subset



	Asteroid	C. F. Gauß	Germany	Franks	Ireland
Asteroid	0	1	1	0	0
C. F. Gauß	1	0	1	1	0
Germany	1	1	0	1	1
Franks	0	1	1	0	1
Ireland	0	0	1	1	0

Figure 2: Illustration of the sliding window mechanism for a window size of 3 and the resulting co-occurrence matrix

3. Results

- Human navigational paths are useful for calculating semantic relatedness between concepts (see Table 1)
- A small portion of human paths (exhibiting a low average mean indegree) shows globally best performance (see Figure 3)

Window size	None	2	3	4	5
Accuracy	0.644	0.636	0.706	0.713	0.686

Table 1: Semantic relatedness accuracy

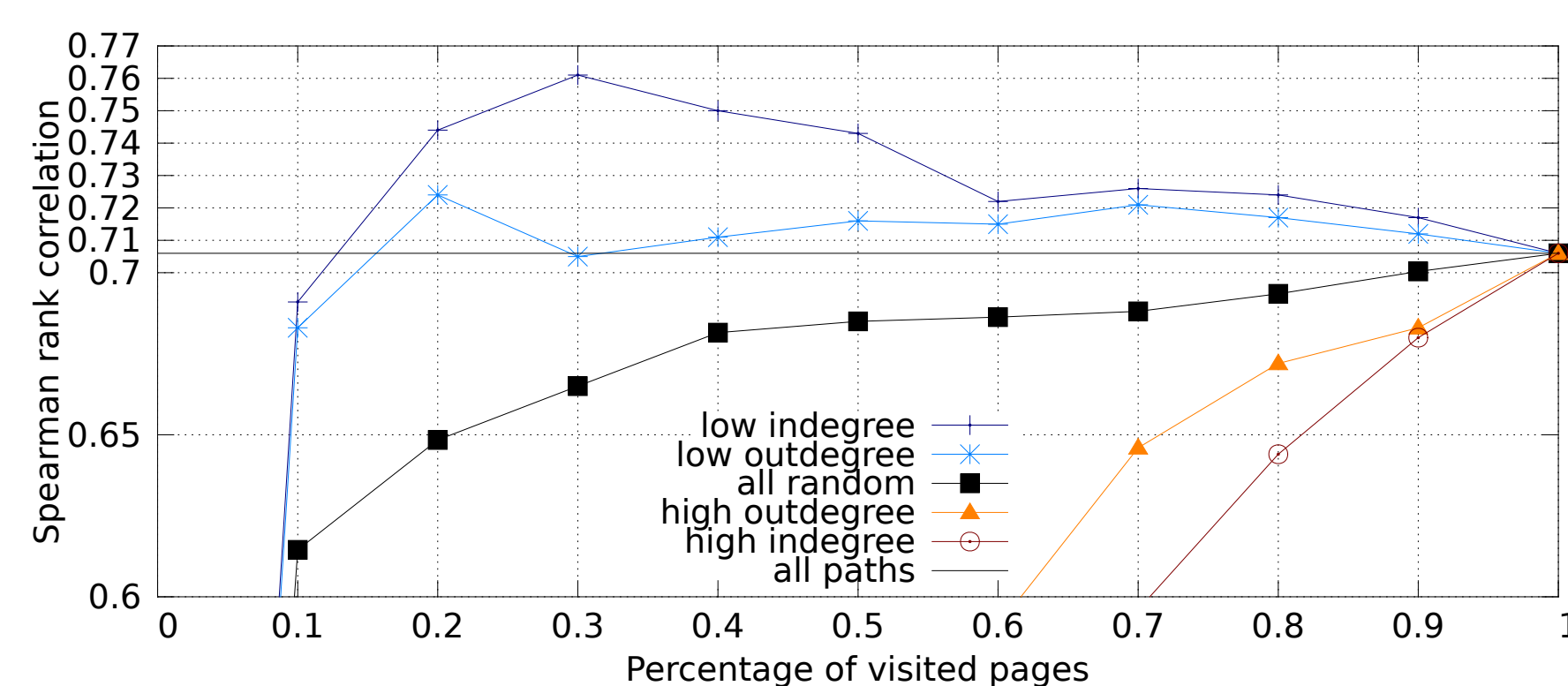


Figure 3: Semantic relatedness of selected path subsets for window size 3

4. Future Work & Outlook

- Extend datasets and path selection strategies
- Investigate implications of our observations on models of human navigation on the Web
- Envision that future methods focus more on usage data
- Produce subjective scores instead of plain objective ones

5. References & Acknowledgements

- [Finkelstein2002] L. Finkelstein, E. Gabrilovich, Y. Matias, E. Rivlin, Z. Solan, G. Wolfman, and E. Ruppín. Placing Search in Context: The Concept Revisited. ACM Transactions on Information Systems, 20(1):116–131, Jan. 2002.
- [West2009] R. West, J. Pineau, and D. Precup. Wikispedia: An Online Game for Inferring Semantic Distances between Concepts. In Proceedings of the 21st International Joint Conference on Artificial Intelligence, IJCAI '09, pages 1598–1603, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.

This work is in part funded by the FWF Austrian Science Fund Grant I677 as well as the DFG through the PoSTS project.