Mapping the Zeitgeist

Brent Hecht[1], Johannes Schöning[2]

[1]Department of Electrical Engineering and Computer Science
Northwestern University
Evanston, IL, USA
brent@u.northwestern.edu[Δ]

[2]Institute for Geoinformatics
University of Münster
Münster, Germany
j.schoening@uni-muenster.de

INTRODUCTION

The concept represented by the term "zeitgeist" is such a universally appealing one that "zeitgeist" is one of the few loanwords adopted into English whole cloth from the modern German language. Literally translated, the term means "ghost of time", but the term is mostly commonly used to represent the idea of the "spirit of the era"[1]. In this paper, we inquire about the spatial component of this spirit. In other words, what is the spatial footprint of an era's zeitgeist? *Where* is this "spirit of the age" hovering in any given era?

Many approaches are possible when attempting to answer these questions. Of course, polling the populace of a community, history experts, and/or world/current events whizzes is one method. However, an automated, general, and more uniform solution is also desirable. Enter GeoSR (Hecht and Raubal 2008), a system that allows users to view the spatial footprint of any concept or entity using the unprecedented quantity and diversity of relationships embedded in the Wikipedia Article Graph (WAG) as well as the concept of semantic relatedness measures from the natural language processing community. While the idea of mapping the zeitgeist is our specific goal in this abstract, the broader objective behind this project was to begin exploration of the analytical possibilities – particularly *quantitative* analysis – of GeoSR's output because we intend GeoSR as a platform for users to easily engage in a wide-range of specialization projects (as well as other pursuits discussed in Hecht and Raubal 2008).

[Δ] Work completed while at the Department of Geography of the University of California, Santa Barbara

[1] http://etext.lib.virginia.edu/cgi-local/DHI/dhi.cgi?id=dv4-74

OVERVIEW OF GEOSR

A detailed overview of GeoSR is provided by Hecht and Raubal (2008). The main purpose of GeoSR is to allow users the ability to geographically explore world knowledge using the relations between entities and/or concepts. Integral to the efficient and effective means of doing so is the application of the first semantic relatedness (SR) measure (Budanitsky and Hirst (2006) provide a good overview of SR research) designed for the Wikipedia Article Graph (WAG). At present, GeoSR functions with the WAGs of 10 different languages, including English, German, Spanish, and French. The first application of this system – that which is used in this abstract – is to the allow the user to input a Wikipedia article and receive a geovisualization of the most semantically related spatial articles (articles with a latitude and longitude included by Wikipedians) to the input article. Critically, GeoSR also allows users to qualitatively see *why* each article is semantically related to the input article in natural language format as shown in Schöning et al. (2008). However, the "why" component is outside the focus of this paper.

METHODOLOGY

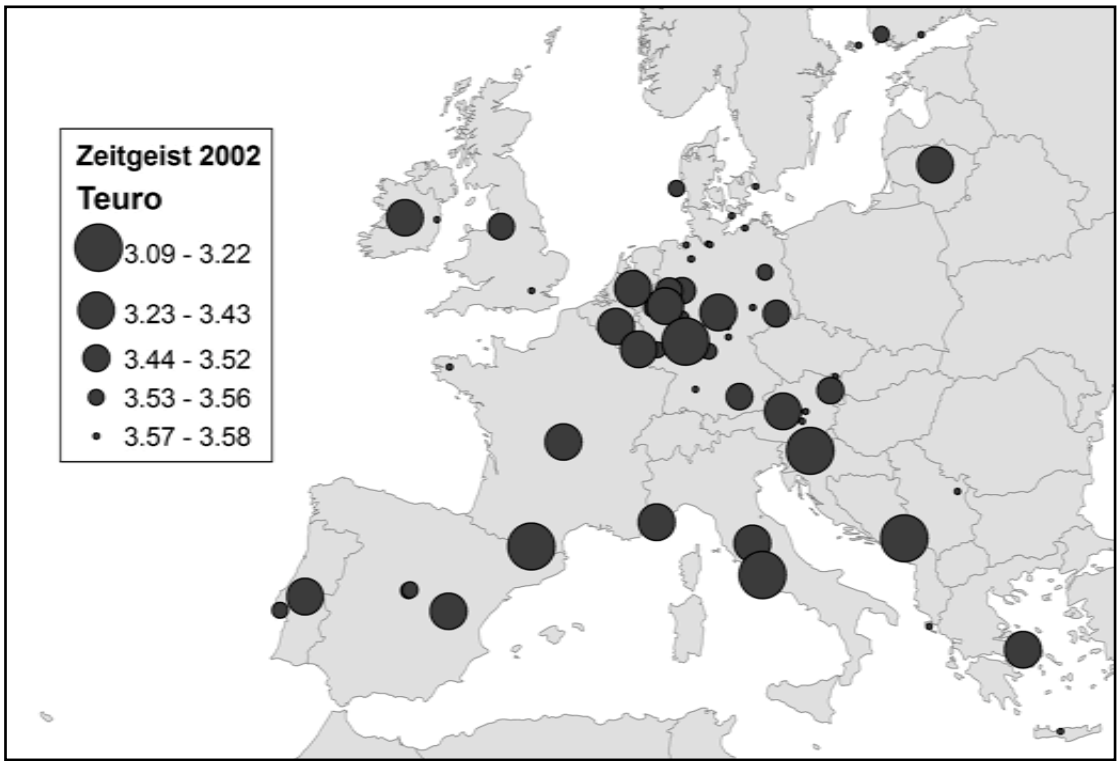
The initial challenge in determining the spatial footprint of a zeitgeist is, of course, defining which concepts and/or entities are part of the zeitgeist in the first place. Although the idea of the "zeitgeist" has seen a huge popularity increase in the past few years thanks to Google's publication of what it calls the annual "Google Zeitgeist"[2], we found Google's definition unappealing, mainly due to its complexity. The Google Zeitgeist is multifaceted and hierarchical; we wanted a simple, single-word definition for each year/era to make this initial analysis straightforward. As such, we turned to the list of "*Wörter des Jahre"* ("words of the year") from the *Gesellschaft für deutsche Sprache* ("Society for the German Language"). This list represents an annual attempt by a group of language experts to capture the spirit of the year in the form of a word that entered common usage during the year (in the German language). Examples include 2002's *"Teuro*" (a play on words that combines the German word for "expensive" and the term "Euro") and 2005's "*Bundeskanzlerin*" (the feminine noun for "Federal Chancellor", which entered the language due to the election of Angela Merkel, Germany's first female federal chancellor). A full list of the *Wörter* (words) used, along with the corresponding pages in the German Wikipedia (which was exclusively used for this analysis) can be found in Table 1.

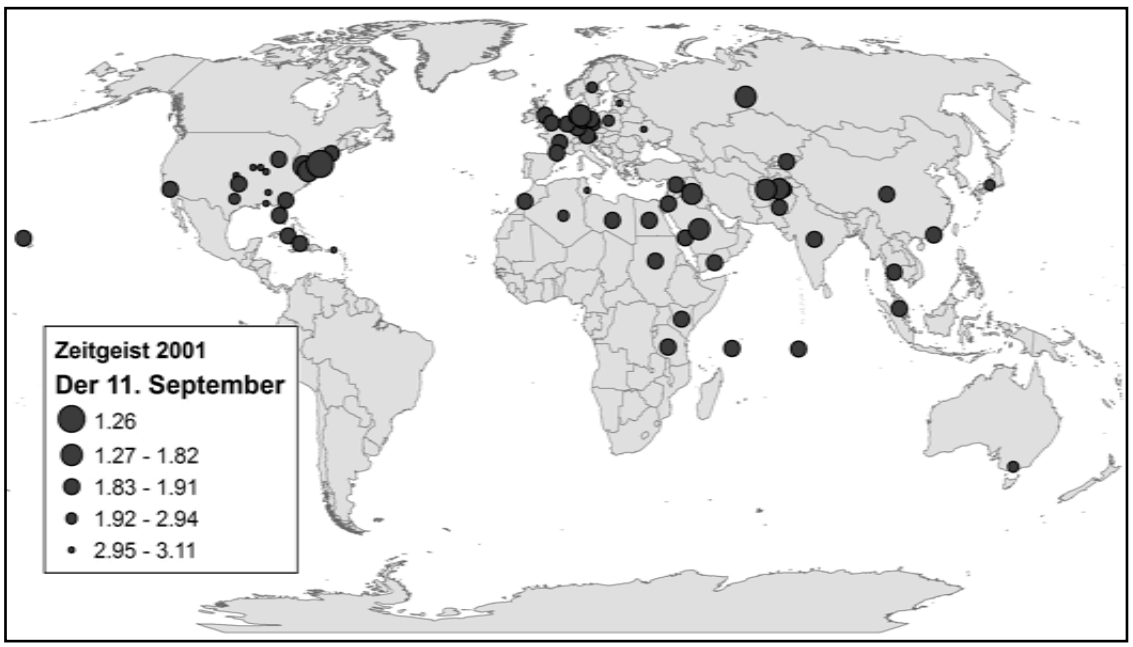
[2] http://www.google.com/intl/en/press/zeitgeist.html

Wort Des Jahre	German Article	Corresponding English Article	Year
Klimakatastrophe	Klimakatastrophe	n/a	2007
Fanmeile	Fanmeile	n/a	2006
Bundeskanzlerin	Angela Merkel	Angela Merkel	2005
Hartz IV	Hartz-Concept*	n/a	2004
das alte Europa	Das alte Europa	Old Europe	2003
Teuro	Teuro	n/a	2002
der 11. September	Terroranschläge am 11. September 2001	September 11, 2001 attacks	2001
Schwarzgeldaffäre	CDU-Spendenaffäre	1999 CDU contributions scandal	2000
Millenium	Jahrtausend	Millennium	1999
Rot-Grün	Rot-Grüne Koalition	Red-green alliance	1998
Reformstau	Reformstau	n/a	1997
Sparpaket	Sparpaket	n/a	1996
Multimedia	Multimedia	Multimedia	1995
Superwahljahr	Superwahljahr	n/a	1994
Sozialabbau	Sozialabbau	n/a	1993
Politikverdrossenheit	Politikverdrossenheit	n/a	1992
Besserwessi	Besserwessi	n/a	1991
Neue Bundesländer	Neue Bundesländer	New Länder	1990
Reisefreiheit	Reisefreiheit	n/a	1989

Table 1: Words of the Year (*"Wörter des Jahre")* used, and their corresponding pages in the German Wikipedia (and English Wikipedia). A German Wikipedia snapshot from September 2007 was used for the analyses. An asteriks indicates that an error occurred while processing the article in this particular snapshot, and thus the word was left out of further analyses.

Each of these Wikipedia articles was then input into GeoSR operating on the German Wikipedia. Shapefiles (ESRI 1998) of the top 100 most semantically related spatial features for each article, as well as their exact GeoSR values, were output by GeoSR. Figure 1 shows four example maps produced using these shapefiles, the collection of which formed the raw data for the quantitative analysis below.


Zeitgeist 2002
Teuro
3.09 - 3.22
3.23 - 3.43
3.44 - 3.52
3.53 - 3.56
3.57 - 3.58


Zeitgeist 2001
Der 11. September
1.26
1.27 - 1.82
1.83 - 1.91
1.92 - 2.94
2.95 - 3.11

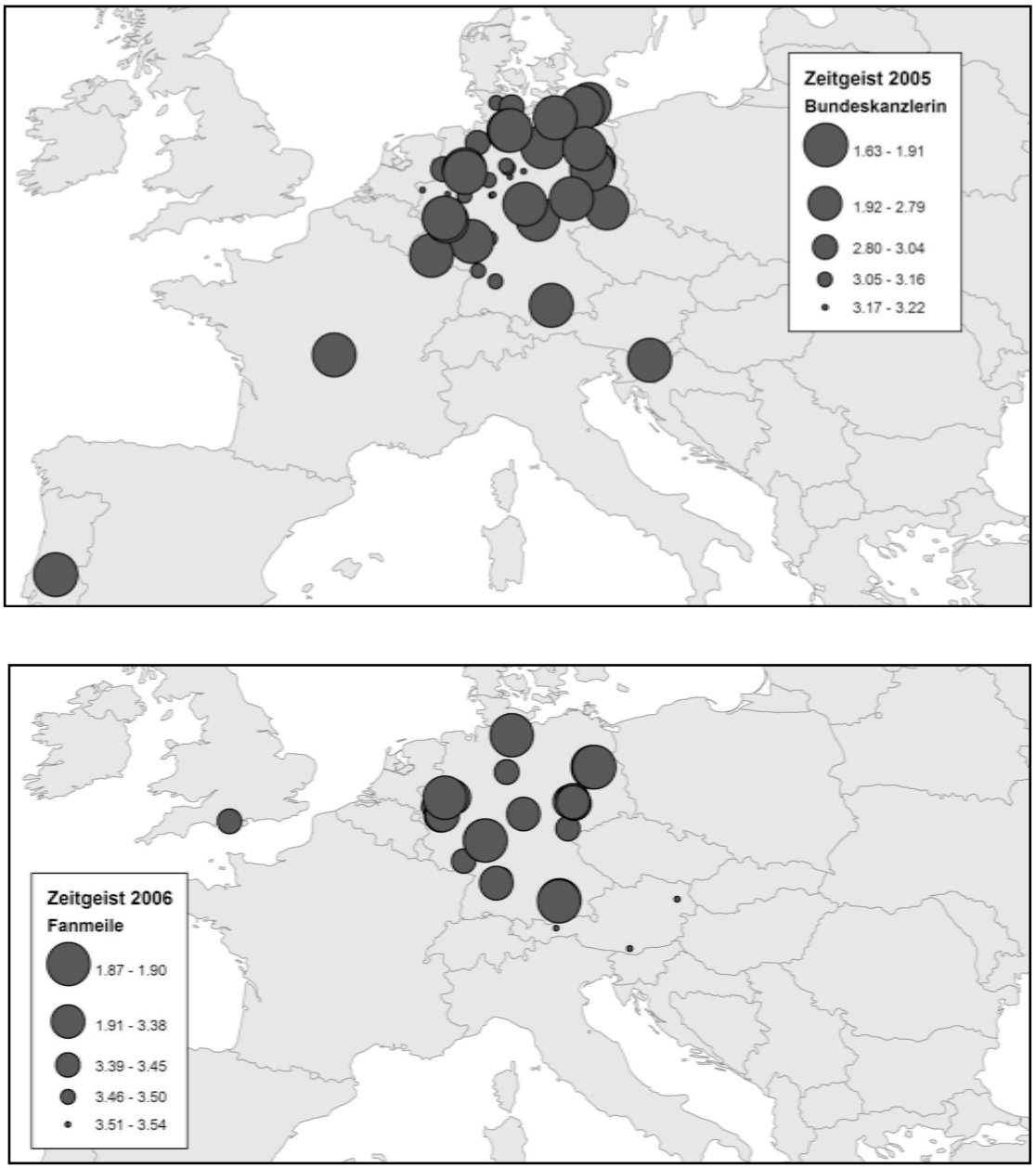


Fig. 1: Visualizations of the top 100 GeoSR values for "Teuro" (2002) (top) and "Terroranschläge am 11. September 2001" (2001) (second), "Bundeskanzerlin" (2005) (third) and "Fanmeile" (2006) (bottom). Note that in the actual implementation of GeoSR, users would be able to click on circles in the visualizations and retrieve a natural language explanation for the size of the cylinder (see Schöning et al. 2008). Also, note that GeoSR values represent semantic distance, the exact inverse of semantic relatedness.

ANALYSES AND RESULTS

The first analysis we performed was to count the percentage of the top 100 features that fell within the borders of modern Germany (hence the start of our study being 1989, the year the beginnings of modern Germany began to form). This analysis proved quite fruitful, as can be seen in Figure 2. Years with high percentages falling within Germany for the most part represent times in which the German people were mostly inward-looking, for instance during the complex and difficult integration of West and East Germany from 1989 to 1991 and in 2006 when the Soccer World Cup took place in Germany and the German team placed third. Conversely, years with high percentages of the most semantically related places falling *outside* of Germany represent the opposite: times in which the German "spirit" was mostly global in focus, for example during the worldwide tumult of 2001 (September 11th) and 2003 (the start of the second Iraq War). As such, we preliminarily conclude that spatial overlay-based analyses with GeoSR data provide interesting results.

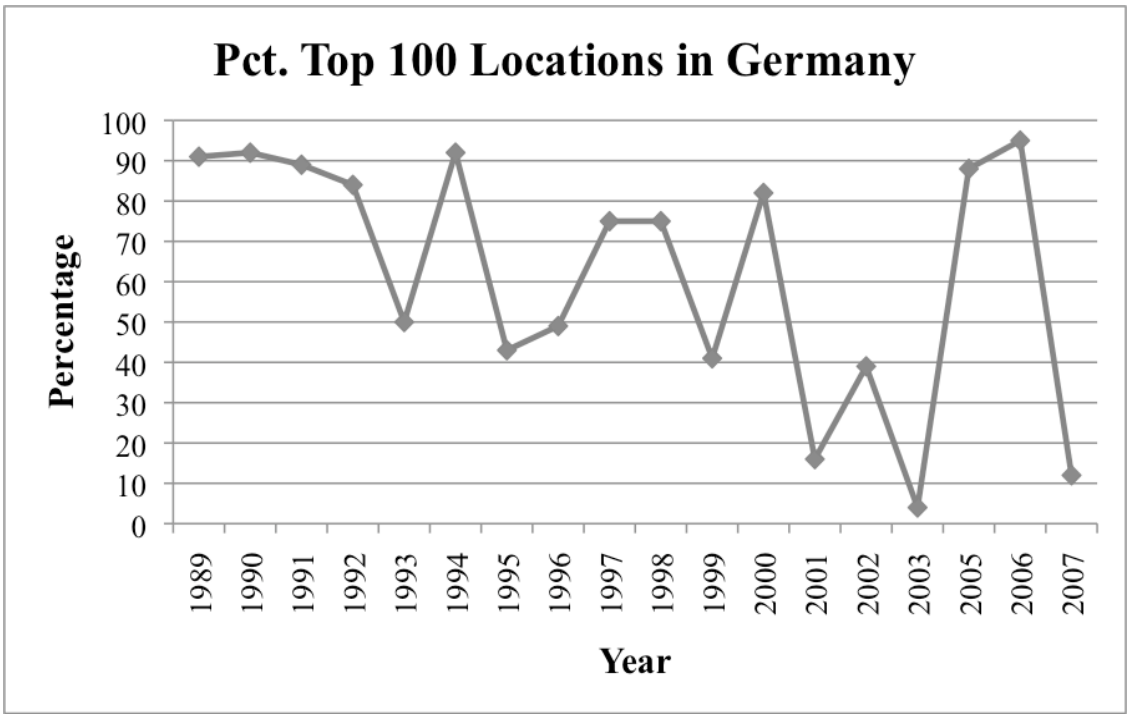


Fig. 2: The results from the first analysis. The chart indicates the percentage of the top 100 most semantic related locations falling within Germany for each word of the year.

The second analysis performed was intended as a measure of the "spatialness" of any given year's zeitgeist. In other words, we evaluated which years' spirits were more directly spatial than others. In theory, this is easily determined using GeoSR, as years in which the top 100 most semantically related locations have a lower average GeoSR value (remember, GeoSR values represent semantic distance, the exact inverse of semantic relatedness) are years in which the "spirit" was more spatially oriented. The results can be seen in Figure 3. It is obvious that three years stick out here: 1996 ("Sparpaket", a package of laws designed to reduce the German budget deficit) has a particularly non-spatial spirit and 2001 ("11. September") and 2005 ("Bundeskanzlerin") have decidedly spatially-oriented spirits. The September 11th article describes an event, an obviously explicitly spatial entity. Many places were involved directly or indirectly in the attacks, and thus the article used to represent the "word" (in this case *words*) directly links to many places many times, and many articles linked to the aforementioned article describe spatial entities (these factors have significant effects on the GeoSR score). The same occurs for the article "Angela Merkel" – which was used to represent the word "Bundeskanzlerin" – because of the biographical (rather than event) nature of the article. 1996's "Sparpaket" potentially represents a problem with this sort of spatial versus non-spatial analysis. The reason the semantic distances are so high in this case is that the "Sparpaket" article is decidedly shorter and contains many fewer inlinks and outlinks than the other articles used in this analysis, which serves to possibly artificially increase the GeoSR score.

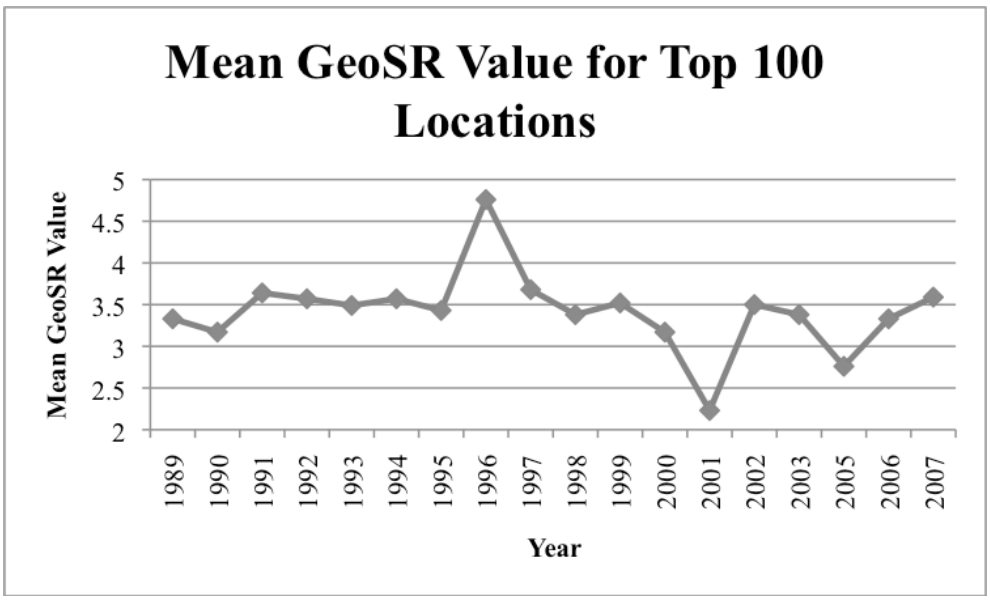


Fig. 3: The results from the second analysis. The chart shows mean GeoSR value for the top 100 locations for each year.

CONCLUSION

In this abstract we have shown that the quantitative component of GeoSR has potential as a platform for high-level spatial analyses not only for the GIScience community, but also for many other disciplines. As we improve GeoSR and its built-in SR measure, we hope to develop and share more advanced analyses. We also hope that users of the system will contribute to this process in a variety of application areas.

ACKNOWLEDGEMENTS

Travel to GIScience has been generously supported by the Oak Ridge National Laboratory and the Collabolab of Northwestern University.

REFERENCES

Budanitsky, A. and G. Hirst (2006). "Evaluating WordNet-based Measures of Lexical Semantic Relatedness." Computational Linguistics **32**(1): 13-47.

ESRI (1998) Shapefile Technical Specification. www.esri.com/library/whitepapers/pdfs/shapefile.pdf

Hecht, B. and M. Raubal (2008). GeoSR: Geographically explore semantic relations in world knowledge. The European Information Society: Taking Geoinformation Science One Step Further. L. Bernard, A. Friis-Christensen and H. Pundt. Berlin, Germany, Springer: 95-114.

Schöning, J., Hecht, B., Raubal, M., Krüger, A., Marsh, M. and Rohs, M. (2008). Improving Interaction with Virtual Globes through Spatial Thinking: Helping users Ask “Why?”. IUI 2008: Proceedings of the International Conference on Intelligent User Interfaces, New York, ACM.