Explanatory Semantic Relatedness and Explicit Spatialization for Exploratory Search

Brent Hecht*, Samuel H. Carton*, Mahmood Quaderi*, Johannes Schöning†, Martin Raubal‡, Darren Gergle*§, Doug Downey*

*Northwestern University Dept. of Electrical Engineering and Computer Science, §Northwestern University Dept. of Communication Studies, †Independent Researcher, ‡ETH Zurich Institute of Cartography and Geoinformation

{brent,sam.carton,quaderi}@u.northwestern.edu, {dgergle,d-downey}@northwestern.edu, mail@johannesschoening.de, mraubal@ethz.ch

ABSTRACT

Exploratory search, in which a user investigates complex concepts, is cumbersome with today's search engines. We present a new exploratory search approach that generates interactive visualizations of query concepts using thematic cartography (e.g. choropleth maps, heat maps). We show how the approach can be applied broadly across both geographic and non-geographic contexts through *explicit spatialization,* a novel method that leverages any figure or diagram – from a periodic table, to a parliamentary seating chart, to a world map – as a spatial search environment. We enable this capability by introducing *explanatory semantic relatedness measures.* These measures extend frequently-used semantic relatedness measures to not only estimate the degree of relatedness between two concepts, but also generate human-readable explanations for their estimates by mining Wikipedia's text, hyperlinks, and category structure. We implement our approach in a system called Atlasify, evaluate its key components, and present several use cases.

Categories and Subject Descriptors

H.3.m [**Information Storage and Retrieval**]: Miscellaneous, H.5.m. [**Information interfaces and presentation (e.g., HCI)**]: Miscellaneous

General Terms

Algorithms, Measurement, Experimentation, Human Factors

Keywords

Semantic relatedness, exploratory search, spatialization, cartography, geography, Wikipedia, text mining, GIScience

1. INTRODUCTION

Exploratory search is an open-ended information seeking activity in which a user aims to better understand a complex concept [39, 40]. While exploratory search has historically accounted for roughly a quarter of Web search query volume [33], it remains challenging using today's search engines due to their focus on closed information requests and navigational queries [40].

In this paper, we leverage thematic cartography's well-known ability to communicate complex geographic distributions [2, 5, 6, 37] to help users understand the complex concepts encountered in exploratory search. While the benefits of cartography are usually limited to geographic inquiries, our approach is made domain-neutral by harnessing general relational knowledge mined from Wikipedia. This means that users can employ thematic cartography to explore concepts not only from a geographic perspective, but also from a chemistry perspective, a politics perspective, a music perspective, or a perspective from any other topic area (even user-defined topic areas).

We have implemented our exploratory search approach in a Web-based system called Atlasify. Given a query concept, Atlasify automatically generates an interactive thematic cartography layer (e.g. a choropleth or heat map) on top of a spatial *reference system* from any domain, such as a periodic table, a U.S. senate seating chart, or a world map. The layer illustrates the degree to which the query concept is related to each spatial entity in the reference system (e.g. chemical elements, senators, countries). By clicking on a spatial entity, users see natural language explanations of exactly how that entity or region is related to the query concept. Users can enter any query that corresponds to a Wikipedia article (i.e. a page title, anchor text, or redirect).

To make this process more concrete, consider the Atlasify use case in Figures 1–4. In Figure 1, a user who wants to learn about nuclear power has queried Atlasify for "Nuclear Power" and selected "Periodic Table" as the desired spatial reference system. As is typical with choropleth maps, the dark green areas in Figure 1 are very related to nuclear power, and the lighter green areas are

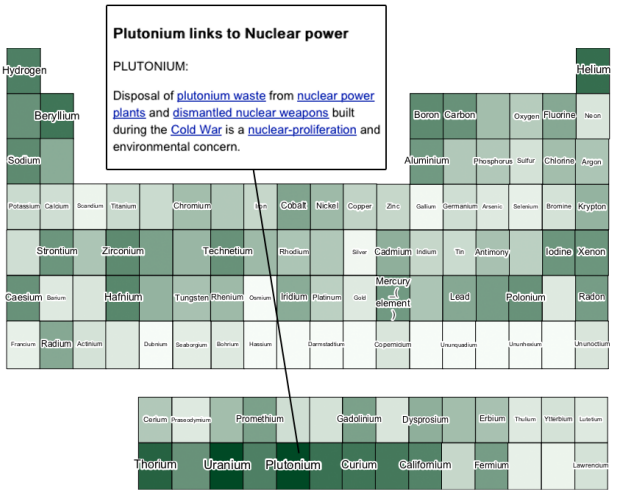

Figure 1. Atlasify's visualization of the query concept "Nuclear Power" on the "Periodic Table" spatial reference system. If users click on plutonium, they receive a list of explanations of how nuclear power and plutonium are related, a list that includes the above explanation.

SIGIR'12, August 12–16, 2012, Portland, Oregon, USA.

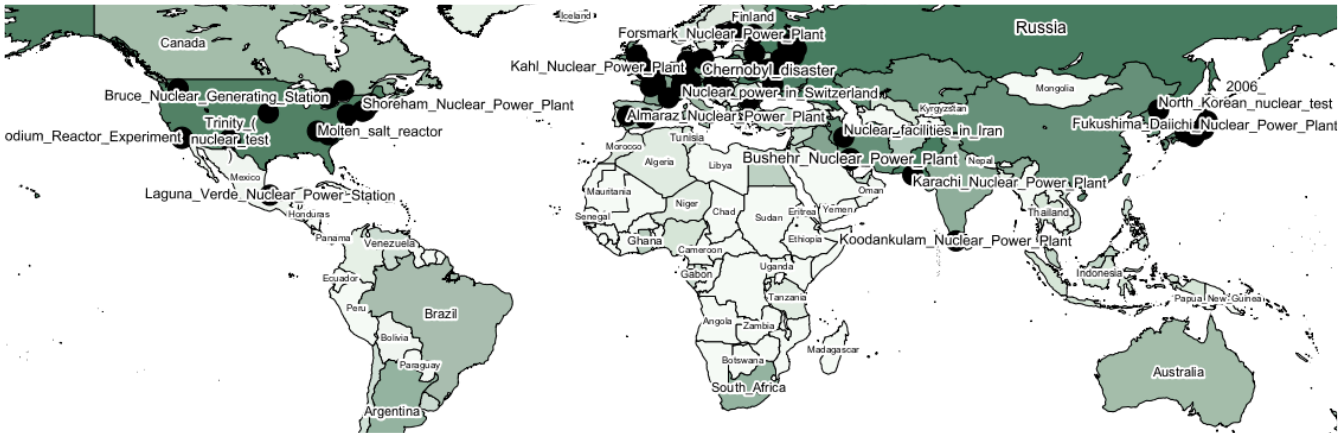

Figure 2. Atlasify visualizing the query concept "Nuclear Power" on the "World Map" reference system. The user is able to see that, for instance, sub-Saharan Africa is not very related to nuclear power, while the United States, Chernobyl, and Fukushima are quite related. The "World Map" reference system is the largest of Atlasify's spatial reference systems, with approx. 380,000 entities. For each query concept, the AtlasifySR+E *semantic relatedness between all entities and the query concept must be calculated.*

less related. Exploring further, the user may wish to understand why, for example, plutonium is so strongly related to nuclear power. By clicking on plutonium in the visualization, the user is presented with natural language explanations of the relationships between nuclear power and plutonium. Seeking a geographic perspective on nuclear power, the user then changes to the "World Map" reference system (Figure 2). The user does the same for a temporal perspective in Figure 3 (the "Timeline" reference system) and a United States politics perspective in Figure 4 (the "U.S. Senate Seating Chart" reference system). Note that Atlasify correctly highlights Fukushima, Russia, and the United States in the world map, the various important eras in the history of nuclear power on the timeline, and so on. While this use case focuses on the query concept "Nuclear Power", Atlasify allows users to query for over 15 million articles from 25 different Wikipedia language editions. These Wikipedia-based concepts can currently be visualized on 13 different reference systems and adding new reference systems is straightforward.

The effectiveness of thematic cartography is well established in geographic domains (e.g. [2, 5, 6, 37]). The goal of our

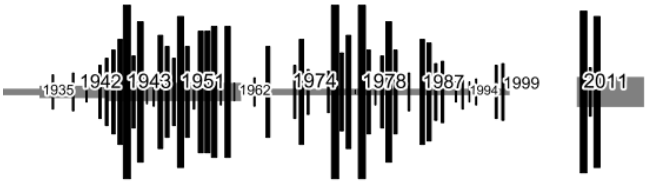

Figure 3. "Nuclear Power" visualized on the "Timeline" reference system.

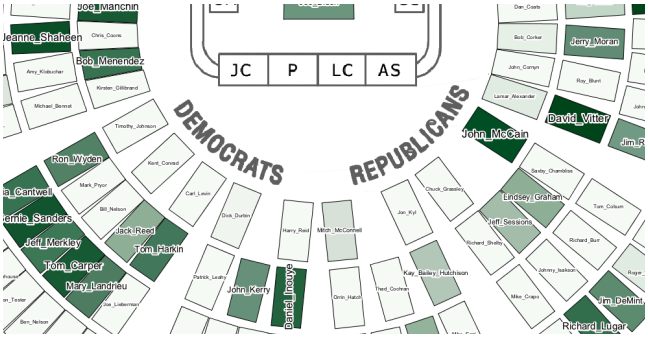

Figure 4. "Nuclear Power" visualized on the "U.S. Senate Seating Chart" reference system.

exploratory search approach is to extend the strengths of thematic cartography to the wide variety of domains and query concepts encountered in exploratory search. This goal can be broken down into three key challenges, the solutions to which are additional contributions of this paper and have implications outside of exploratory search.

The first challenge involves generalizing the visualization strategy used in Figure 2 to non-geographic reference systems (e.g. periodic tables, anatomical charts, timelines, and many other figures and diagrams). Our solution is *explicit spatialization* (ES), which enables cartographic and geographic information retrieval (GIR) methods to be applied in any figure or diagram. As discussed in Section 3, ES accomplishes this by "spatializing" concepts into pre-defined reference systems and generalizing the canonical model of geographic information to incorporate domain-neutral spatial information. In doing so, ES can extend the ongoing advances in online mapping and GIR to many domains outside of geography.

The second challenge involves automatically estimating the degree of relatedness between any of the millions of possible query concepts (e.g. "Nuclear Power") and every spatial entity in each reference system (e.g. chemical elements, countries). These estimates determine the value of the visual variables manipulated in thematic cartography, such as color and text size (e.g. the shades of green and font sizes in Figures 1–4). We show how Wikipedia-based semantic relatedness (SR) measures, which provide a numerical relatedness score for any pair of lexically expressed concepts, can solve this problem. We introduce a new SR measure, *AtlasifySR+E,* which uses a learned model to combine six separate SR measures, each capturing a different type of relationship. Experiments on several SR benchmarks show that *AtlasifySR+E* achieves state-of-the-art performance while also remaining language-neutral and using only open, easily accessible data, overcoming two limitations of the current state-of-the-art SR measure.

The final challenge concerns generating natural language explanations of the relationships between the query concept and any spatial entity. These explanations realize the key paradigm of modern interactive cartography that users be able to click on a part of a map to obtain additional details [34, 35]. To address this challenge, we introduce the notion of *explanatory semantic relatedness measures* (SR+E), which not only return a numeric estimate of the semantic relatedness between two entities, but also

explain the identified relationships to end users. We show how Wikipedia-based SR measures can be made explanatory by using machine learning to mine informative snippets of Wikipedia text. Furthermore, we describe how our SR+E measure, *AtlasifySR+E*, uses machine learning to combine the explanations of its six constituent measures. Again, the approach of integrating the perspectives of each SR measure results in improved performance: our experiments demonstrate that *AtlasifySR+E*'s explanations outperform those of any single measure and other baselines.

In summary, this paper presents both a method for leveraging thematic cartography for domain-neutral exploratory search and the innovations in SR and information spatialization required to make that possible, namely (1) explicit spatialization, (2) improved SR estimation, and (3) explanatory SR measures. The remainder of this paper is organized as follows. After covering related work, we formally define explicit spatialization and discuss how it is implemented in the Atlasify system. Next, we introduce explanatory SR measures and describe methods for generating explanations for six separate SR measures. We then present our experiments, in which our approach and its components are evaluated up to the point where thematic cartography's well-understood methods take over. Finally, we conclude and discuss directions for future work.

2. RELATED WORK

In this section, we cover research related to this paper at a high level, with additional related work specific to each section of the paper discussed in context. Our research falls into the area of exploratory search. White et al. write that exploratory search systems aid users with information seeking problems that are "open-ended, persistent and multi-faceted" [40]. This stands in contrast to traditional Web search, which is primarily concerned with navigational queries and closed information requests. Despite the prevalence of exploratory queries, exploratory search is a relatively new research area with many open questions [40].

The field of cartography has identified several reasons why humans find thematic mapping useful for understanding complex geographic patterns. The known benefits of thematic maps are the communication of specific information [20, 37], the communication of regional/general information [20, 37], straightforward comparisons between maps showing different distributions [37], and straightforward comparisons between a mapped distribution and one's mental model of depicted entities and regions [24, 37]. We enable these benefits in a wide variety of domains outside geography. For instance, in Figure 1 it is easy to see that plutonium specifically is quite related to nuclear power, but so is the entire "region" of actinides (the bottom row). An Atlasify user may recall from chemistry class that actinides have to do with the atomic age, so the fact that this region is highlighted reinforces her mental model. Finally, comparing Figure 1 with a periodic table visualization of, say, coal power, it is easy to identify differences in the chemistry of the two concepts.

Our work within geographic reference systems is related to research in language models associated with geographic places For example, Google Correlate [12] provides an interface to models based on georeferenced query logs. Others have leveraged geographic language models to study the geographic distribution of *zeitgeist* terms [16], to explore the use of relatedness-like metrics in a geographic context [15, 28], and for various other applications (e.g. [7,19,27]). Some of this research has been echoed in the temporal domain (e.g. [28, 30]). We extend this work by generalizing the notion of geographic language models to arbitrary spatial reference systems, rather than just geographic and temporal ones. This research is also the first to our knowledge to (1) use geographic language models for exploratory search, (2) apply robust SR measures to geographic language models, and (3) use explanatory SR measures in this context (or any other).

3. EXPLICIT SPATIALIZATION

Explicit spatialization (ES) is a novel form of information spatialization that, diverging from the existing spatialization literature, uses *pre-defined* reference systems (e.g. maps, figures, and diagrams) instead of *data-driven* reference systems. While ES is essential to our exploratory search approach, it also has implications beyond this work. Namely, it provides a new means by which advances in online mapping and geographic information retrieval (GIR) can be extended to domains outside of geography.

3.1 Definition of Explicit Spatialization

Explicit spatialization (ES) "spatializes" or "projects" any object *o* into a pre-defined reference system such as a periodic table, map, or seating chart. More formally, ES defines a process that represents an object *o* in terms of the spatial entities *E* in a reference system *rs* according to the output of an ES function $f_{ES}(o, E)$. We clarify the key elements of this process below.

Let us consider Atlasify's implementation of explicit spatialization. In Atlasify, each object *o* is a query concept (e.g. "Nuclear Power") and the system's ES function is our SR+E measure *AtlasifySR+E*. The spatial entities considered include countries (and cities, landmarks, etc.) in the "World Map" reference system, chemical elements in the "Periodic Table" reference system, and so on. Atlasify therefore spatializes each query concept into each reference system by running *AtlasifySR+E* on each query concept/spatial entity pair.

In explicit spatialization, each spatial entity $e \in E$ in a reference system *rs* is comprised of a tuple <**x**,**d**>, where **x** is a location (spatial footprint) in *rs*, and **d** is one or more data resources describing the entity. These data resources are mined by the ES function to spatialize the object *o*. In Atlasify, **d** consists of a single Wikipedia article describing each spatial entity.

The output of an ES function is a spatial distribution ("layer") whose data model is a generalization of the canonical model of geographic information [11] (see Figure 5). The canonical geographic model formalizes an atomic unit of geographic information as a tuple <**x**, **z**>, where **x** is a location in space-time of an entity on or near the surface of the Earth (e.g. its latitude / longitude coordinate or its polygonal representation) and **z** is a set of *attributes* corresponding to that entity (e.g. temperatures,

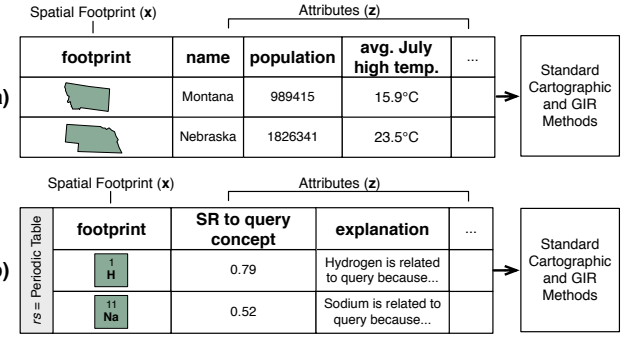

Figure 5. An example of the canonical data model of geographic information (a, top) and the explicit spatialization data model (b, bottom).

population counts). ES generalizes the geographic information model by replacing "the Earth" with an arbitrary reference system *rs* (such as the periodic table, an anatomical chart, etc.). The new model is equivalent to the geographic information model for a single fixed *rs* (except, of course, for the domain of the entities). It is via this reduction that ES can use traditional cartographic and GIR methods with little to no modification.

The flexibility of the ES data model makes it adaptable to nearly any reference system in any domain. As one example of ES's generality, consider a Web browser reference system that, as a user browses the web, shows heat maps visualizing relatedness to a persistent concept of interest. We have implemented a static proof-of-concept of this idea in Atlasify's "New York Times Homepage" reference system (Figure 6).

It is important to note that Atlasify's implementation of explicit spatialization is far from the only possible approach. Other ES functions could include topic detection techniques or an algorithm that calculates entity-level sentiment. Similarly, a projected object *o* could be a blog post, an academic paper, or even an entire document collection, and data resources **d** considered for each spatial entity could include tweets, images, or photo tags.

3.2 Relationship to Traditional Spatialization

Traditional spatialization produces data-driven, abstract reference systems generally by applying dimension reduction to document collections for the purpose of visualizing those collections (see [13, 36] for an overview). ES, on the other hand, leverages *existing* reference systems (e.g. the periodic table, the human body, the surface of the Earth) for general IR applications. As a result, ES avoids the pitfalls that can make traditional spatialization undesirable for search [13], such as the error introduced by dimension reduction [13, 17, 36] and the imposition of a single, static visualization for an entire document collection [13]. While a few commercial document visualization systems (e.g. [32]) have begun to explore extensible reference systems as in ES, they still rely on traditional spatialization as their primary paradigm.

3.3 Spatiotagging

Preparing a new reference system for an ES application like Atlasify is a straightforward process that we call *spatiotagging*. Spatiotagging is a generalization of geotagging to arbitrary spatial reference systems. To construct a reference system using spatiotagging, one simply identifies the spatial footprint (**x**) of the spatial entities in the reference system (e.g. chemical elements, Senate chamber seats, countries), and matches those entities to data resources (**d**) (e.g. corresponding Wikipedia articles). Spatial footprints can be identified by manually tracing the shapes of entities over a figure, diagram or image, obtaining pre-existing spatial representations (e.g. KML files or shapefiles), leveraging computer vision (e.g. OCR), or utilizing other techniques.

3.4 User-defined Reference Systems

While spatiotagging a new reference system is straightforward, it requires some effort. Further, users may not be able to find an existing reference system appropriate for their needs. In this section, we show how it is possible to extend ES to support ad-hoc, user-defined reference systems via the semantic relatedness

Figure 6. Atlasify's explicit spatialization of the query concept "Arab People" on the "New York Times Homepage" reference system. Atlasify correctly understands that the left column consists of Arab people-related stories. It also detects small increases in relatedness near "crude oil", etc.

work explained in detail later in the paper (Sections 4–5).

Explicit spatialization enables the automatic construction of user-defined reference systems through three components: (1) predefined templates, which describe the general layout of the reference systems, (2) a category of concepts to act as spatial entities, and (3) SR algorithms. Figures 7 and 8 provide a small use case of user-defined reference systems generated in Atlasify using the "spectrum" and "simplex" predefined templates respectively. In both figures, the spatial entity concepts are members of the Wikipedia category "Grammy Award winners," and the SR algorithm is *AtlasifySR+E*.

Predefined templates and their "anchor concepts" make user-defined reference systems explicit. The "spectrum" template supports two anchor concepts and the "simplex" supports three. In Atlasify's implementation, users can set these anchor concepts to any concept covered by a Wikipedia article (e.g. "Rock music", "Hip hop music"). Note that a reference system defined by a given set of anchor concepts remains fixed, independent of which category of concepts (spatial entities) or query concept is plotted on it (i.e. it is not data-driven). As noted above, this is the key distinction between explicit and traditional spatialization.

In user-defined reference systems, the exact position of each spatial entity is defined by the SR between the corresponding concept and each of the anchor concepts. In other words, spatiotagging is done automatically in these reference systems using SR. If a spatial entity is very close to an anchor, this indicates that the corresponding concept is significantly more related to the nearby anchor than to the others. Only concepts that are non-trivially related to all anchors are included as spatial entities. In Atlasify's implementation, the category of concepts to act as spatial entities can be any Wikipedia category (Section 4).

As shown in Figure 8, user-defined reference systems are intended to be used as the basis for thematic cartography visualizations of query concepts just like standard ES reference systems. However, user-defined reference systems may also have value as exploratory search tools in and of themselves, without the thematic layer (e.g. Figure 7), but this more closely resembles traditional spatialization.

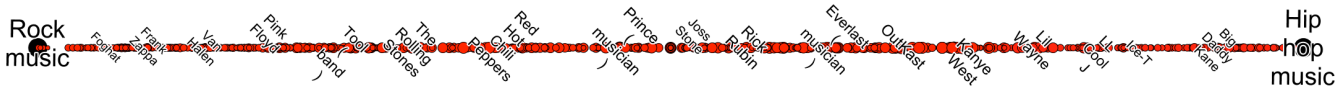

Figure 7: A spectrum user-defined reference system of Grammy Award winners plotted from "Rock music" to "Hip hop music". When used for thematic cartography visualizations, relatedness to the query concept is displayed in a similar fashion to Figure 3.

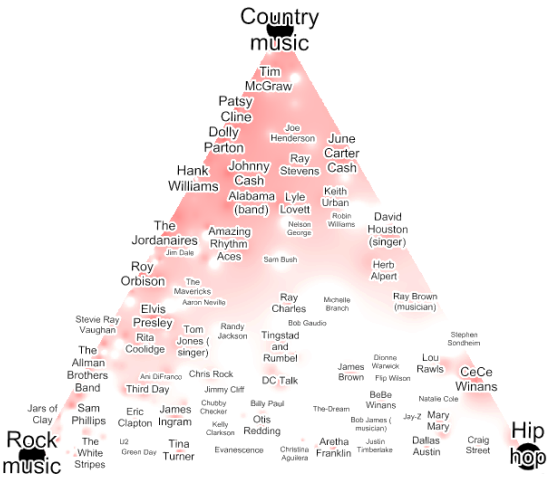

Figure 8. The query concept "Grand Ole Opry" is visualized on a simplex reference system defined by music genres, with the spatial entities being members of the Wikipedia category "Grammy Award Winners". It is clear that "Grand Ole Opry" is more related to country music than, say, rock music.

3.5 Spatial Information Retrieval

Our exploratory search approach focuses on the cartographic benefits of explicit spatialization, but ES also has GIR implications. Namely, ES generalizes GIR to *spatial information retrieval*[2] (SIR). In SIR, many GIR research areas – from understanding vague regions to toponym (place name) resolution to geographic relevance ranking to local search – can become relevant in non-geographic domains. For instance, Jones et al.'s work on modeling vague geographic regions [19] like the English Midlands could be applied to numerous other reference systems, e.g. to model the "belly" or "tummy" vague regions in an anatomical reference system.

To demonstrate the possibilities of SIR, we have implemented in Atlasify one of the most basic GIR features: the simple bounding box spatial query. Users can issue these spatial queries by clicking Atlasify's "What's Related Here" button. Users are then presented with a list of concepts ranked by relatedness to the spatial region defined by the current view frame, which can then be filtered by Wikipedia category. This allows users to, for example, find out the concepts most related to the actinide elements or to the longest-serving members of the Democratic caucus (in the middle left of the seating chart).

4. EXPLANATORY SEMANTIC RELATEDNESS MEASURES

In this section, we introduce *explanatory semantic relatedness measures* (SR+E). Like traditional semantic relatedness (SR) measures, SR+E measures return a value that summarizes the number and strength of relationships between a given pair of concepts (e.g. <Nuclear Power, Plutonium>)[14]. However, along with each value, SR+E measures also provide a ranked list of natural language *explanations* of the various relationships underlying the value, in descending order of informativeness.

As noted above, SR+E measures play an integral role in our exploratory search approach. Each spatial distribution that our approach visualizes with thematic cartography is made up of SR estimates between the query concept and all entities in a reference system. While these visualizations show users the degree to which a query concept is related to a given spatial entity, the natural language explanations produced by SR+E measures describe *why* they are related. In doing so, the explanations provide users with "details-on-demand" [35] for a clicked spatial entity, following the principles of interactive cartography [34].

We begin our detailed discussion of SR+E measures below by introducing methods for generating explanations for three popular existing SR measures – *WikiRelate* [38], *MilneWitten* [22], and *Explicit Semantic Analysis* [9] – and then do the same for several new measures. We describe how each resulting SR+E measure mines Wikipedia's text, links, or category structure to create explanations that reflect the relationships captured by the corresponding SR measure. We next cover *AtlasifySR+E*, the SR+E measure used in our exploratory search approach and implemented in the Atlasify system. *AtlasifySR+E* combines the benefits of the individual SR+E measures discussed above using a learned model. We hypothesized this ensemble approach could produce better SR estimates and explanations than any single measure alone. While we discuss the design of *AtlasifySR+E* here, our evidence that supports this hypothesis and descriptions of the related machine learning experiments are in Section 5.

Finally, we note that while our focus in this paper is on utilizing SR+E for exploratory search, we expect the explanation mechanisms and improved SR measures will have broader applicability as well. SR estimates are frequently utilized in NLP, AI, and IR [4, 9, 43], and have been applied in tasks such as information extraction [4], clustering [1, 30] and search [30].

4.1 Adding Explanations to SR Measures

4.1.1 Selecting SR Measures

There are many SR measures in the literature. Even limiting our attention to Wikipedia-specific SR measures, which have been shown to be better [30] or as good as WordNet-based measures [43], there are still quite a few to consider (e.g. [9, 22, 30, 38]). We focus on three such measures – *WikiRelate*, *MilneWitten*, and *Explicit Semantic Analysis* – because they are among the best-known SR measures and because each uses a different Wikipedia lexical semantic resource [44], thereby capturing different types of relationships between concepts. We also introduce new SR measures to take advantage of additional types of relationships that are not identified by published SR measures.

4.1.2 General Approach to Adding Explanations

Before discussing the details of how we added explanatory capabilities to each of the individual SR measures, we first cover the elements that apply across all measures.

As noted above, each SR+E measure must return a list of natural language relationship explanations ranked by informativeness. While there are other possible approaches, here we define the *informativeness* of each explanation to be based on two factors: the strength of the described relationship and the quality of the textual description. As such, each explanation must consist of natural language text, a relationship strength value, and a text quality value.

In all of our SR+E implementations, the text of relationship explanations is mined from Wikipedia. Several of the SR measures we considered implicitly calculate relationship strength when computing SR values. Where this is not true, we have developed strength metrics that are consistent with the SR measure's overall algorithm. As is described in Section 5, we

[2] The term "Spatial Information Retrieval" has been used as a synonym of GIR in the literature, but not, to our knowledge, to refer to the generalization of GIR as it is used here.

utilize machine learning techniques to map features of the textual explanations to an estimate of text quality, and combine this with relationship strength using heuristics to arrive at a final ranking. The heuristics differ for each SR+E measure, but they generally weigh relationship strength more heavily than text quality.

4.1.3 WikiRelate

WikiRelate [38] uses the Wikipedia Category Graph (WCG) structure as its lexical resource. In Wikipedia, each article can have 0 or more categories (they appear at the bottom of the article), and each of these categories can have 0 or more parents. The resulting graph is the WCG, which is a *folksonomy* [38].

WikiRelate leverages a variant of the WCG path length between the articles *a* and *b* to estimate SR(*a*,*b*). The insight behind this design is that each path represents a relationship between *a* and *b*, and the shorter the path, the stronger the relationship. We construct *WikiRelate* explanations to elucidate these relationship paths to the user in natural language. For example, Figure 9 displays a *WikiRelate* explanation for the strongest relationship between Chemistry and Mathematics (the shortest WCG path between the two articles). In the case of *WikiRelate*, text quality is not considered in the informativeness function as the natural language is automatically determined in the same way for all explanations.

Chemistry and Mathematics both belong to Category:Academic Disciplines

Chemistry

Chemistry is a member of Category:Natural sciences, which is a member of Category:Academic disciplines

Mathematics

Mathematics is a member of Category:Formal sciences, which is a member of Category:Academic disciplines

Figure 9. The top WikiRelate *explanation for the concept pair <Chemistry, Mathematics>. The format of this explanation and those in Figures 10 and 11 mimic that of the Atlasify interface.*

4.1.4 MilneWitten, OutlinkOverlap, and WAGDirect

The Wikipedia Article Graph (WAG) consists of Wikipedia articles and the links between them. There are several published WAG-based SR measures (e.g. [15, 22]). We implemented the measure by Milne and Witten [22], *MilneWitten*, as it has been used in popular web mining applications (e.g. [23]).

MilneWitten operates by comparing the set of articles that link to the articles *a* and *b*. The intuition is that if *a* and *b* share many inlinks, they should be assigned a high SR score.[3] The relationships considered here are indirect: a shared inlink means that an article *c* links to both *a* and *b*. Explanations based on *MilneWitten* must therefore elucidate the nature of these $a \leftarrow c \rightarrow b$ relationships. Figure 10a displays the most informative *MilneWitten* explanation for the concept pair <United States, Chocolate> (*c* = "Chocolate chip").

However, in order to establish that the explanation in Figure 10a was the top explanation – recall that explanations are ranked by informativeness, which is a function of strength and text quality – our *MilneWitten+E* implementation needed a way to measure the strength of each $a \leftarrow c \rightarrow b$ relationship. In other words, we required some method of determining that "Chocolate chip" represents a stronger $a \leftarrow c \rightarrow b$ relationship than, say, the article "List of Viva Piñata Episodes", which also links to both "Chocolate" and "United States". To solve this problem, we use bootstrapping to calculate $SR_{MilneWitten}(a,c)$ and $SR_{MilneWitten}(b,c)$. The strength of each $a \leftarrow c \rightarrow b$ relationship is then computed by taking $SR_{MilneWitten}(a,c) * SR_{MilneWitten}(b,c)$. This algorithm results in the relationship involving "Chocolate chip" being deemed the strongest relationship, with that involving "List of Viva Piñata Episodes" much further down the list.

Chocolate chip links to United States and Chocolate

CHOCOLATE CHIP:

Chocolate chips are small chunks of chocolate.

CHOCOLATE CHIP: AVAILABILITY

Today, chocolate chips are very popular as a baking ingredient in the United States and the chocolate chip cookie is regarded as a quintessential American dessert.

Cheese links to France

CHEESE: WORLD PRODUCTION AND CONSUMPTION:

The biggest exporter of cheese, by monetary value, is France; the second, Germany (although it is first by quantity)

Life and Death link to Organism

LIFE:

In biology, the science of living organisms, life is the condition that distinguishes active organisms from inorganic matter. Living organisms undergo metabolism, maintain homeostasis, possess a capacity to grow, respond to stimuli, reproduce and, through natural selection, adapt to their environment in successive generations.

DEATH:

Death is the termination of the biological functions that sustain a living organism.

Figure 10: (a, top) The top MilneWitten *explanation for <United States, Chocolate>, (b, middle) the top* WAGDirect *explanation for <Cheese, France>, (c, bottom) the top* OutlinkOverlap *explanation for <Life, Death>.*

We have also implemented a modified version of *MilneWitten*, *WeightedMW,* that more heavily weights the links that occur at the very beginning of overlapping articles (i.e. in the *gloss* of the article). The experiments in Section 5 show that this weighted measure estimates SR values somewhat better than our implementation of *MilneWitten*. Explanations are generated in the same fashion as in standard *MilneWitten*.

MilneWitten and *WeightedMW* ignore two important types of relationships present in the WAG. First and foremost, if *a* links directly to *b* ($a \rightarrow b$) or vice versa ($b \rightarrow a$), this link obviously represents a significant relationship between *a* and *b*. We implemented a new SR measure called *WAGDirect* to capture these relationships. *WAGDirect* considers only direct links between *a* and *b*, weighted by whether the link occurs in the gloss of the article and whether the link is bidirectional. Explanations of *WAGDirect* relationships thus consist of text snippets from article *a* that discuss *b*, and/or vice versa (Figure 10b), without any intermediary article *c*.

It was also important that we consider the inverse of *MilneWitten*: the overlap of the set of outlinks of *a* and *b*. This is done by our *OutlinkOverlap* measure, which is similar to other algorithms explored by Milne and Witten [21]. *OutlinkOverlap* uses the principle that, broadly speaking, if *a* and *b* share a significant number of outlinks, then *a* and *b* are quite related. *OutlinkOverlap* explanations thus describe how *a* and *b* discuss these mutually outlinked articles. In other words, they include text snippets from *a* and *b* that explicate the $a \rightarrow c \leftarrow b$ relationships considered by this SR measure (see Figure 10c). *OutlinkOverlap* relationship

[3] Our implementation of *MilneWitten* is slightly simplified from Milne and Witten's final measure; we only consider their "Google Distance-inspired" metric. They were able to gain modest but insignificant improvements by averaging in their "TFIDF-inspired" metric.

strengths are calculated in a similar manner as *MilneWitten* strengths.

4.1.5 Explicit Semantic Analysis

Explicit Semantic Analysis (*ESA*) [9,10] is a popular SR algorithm that uses Wikipedia text as its lexical resource. *ESA* models input concepts *a* and *b* "in terms of Wikipedia-based concepts" [9]. The measure is "explicit" because Wikipedia articles, which are understandable to humans, define this modeling space. *ESA*'s use of real concepts stands in stark contrast to the abstract concepts of methods like Latent Semantic Analysis (LSA), just as the real spaces of explicit spatialization differ from the abstract spaces of traditional spatialization (*Explicit Semantic Analysis* motivated the name of explicit spatialization).

Broadly speaking, to produce SR estimates, *ESA* considers the co-occurrence of *a* and *b* in a large number of Wikipedia articles *C*. Specifically, *ESA* represents *a* and *b* as vectors of bag-of-words similarity to each article *c* in *C*. It then compares these vectors using cosine similarity. The relationships considered by *ESA* are thus co-occurring mentions of *a* and *b* in each Wikipedia article in the concept space. Stronger relationships are defined by articles in *C* that more frequently mention both *a* and *b* (with consideration for document frequency as well), and strength can be estimated by comparing the combined values in each vector dimension while calculating the cosine similarity. Explanations derived from ESA thus describe the co-occurrence of mentions of *a* and *b* in each article *c* in *C* in a human-readable fashion (Figure 11).

Beer in Ireland discusses both Ireland and Beer

Though Ireland is better known for stout, 63% of the beer sold in the country is lager.

Figure 11: The top ESA *explanation for <Ireland, Beer>. In this case c = "Beer in Ireland".*

4.2 AtlasifySR+E

The SR+E measures discussed above capture distinct relationship types. *WikiRelate* tends to operate on classical relations [4] such as *isA* (hyponymy/hypernymy) and *hasA* (meronymy/holonymy) [38]. The WAG-based SR measures are more capable of discovering non-classical relations [4], such as *isTheBiggestExporterOf* (Figure 10b). Finally, *ESA* discovers the "distributional" relationships [4] inherent to text co-occurrence.

AtlasifySR+E, the algorithm employed in our exploratory search approach, combines all six previously discussed SR measures. The goal in doing so was to develop an SR+E measure that understands all three types of relationships. We hypothesized that such an ensemble measure would produce both (1) better SR estimates and (2) better relationship explanations. *AtlasifySR+E's* SR estimate for a pair of terms is the output of a learned model whose features include the estimates of each constituent SR measure as well as features like the word sense entropy of the pair. *AtlasifySR+E* generates explanations for these estimates using a different learned model to select the best explanation among those output by each constituent measure. *AtlasifySR+E* then iterates, choosing the next best explanation, resulting in a long ranked list of explanations.

The experiments section that follows (Section 5) describes in detail each of the learned models and their associated machine learning experiments. We also show in Section 5 that both of our hypotheses related to the combining of SR measures for improved performance were supported.

5. EVALUATION EXPERIMENTS

Evaluation of exploratory search systems is a notoriously difficult problem [39, 40]. In this paper, our evaluation strategy is to investigate the performance of the individual components of our exploratory search approach. Specifically, we focus the evaluation on our method of projecting query concepts into spatial distributions using SR+E's relatedness estimates and explanations. This has the added value of confirming these components as independent contributions. Once these spatial distributions have been created, thematic cartography's well-evaluated techniques (see [37] for an overview) can be employed.

Below, we first describe experiments that demonstrate the state-of-the-art accuracy of our SR estimates. Next, we discuss how we collected over 2,500 human judgments of explanation quality and used these judgments to train a ranker whose performance significantly exceeds baseline approaches.

5.1 SR Value Experiments

Accurate SR value estimates are integral to our exploratory search approach. The colors, text sizes, and other visual variables in Figures 1–4 and 7–8 are defined by *AtlasifySR+E*'s estimates of the SR between each spatial entity and the query concept. Our method for achieving high-quality SR is to combine the estimates of the six SR measures mentioned above using machine learning, and use the resultant trained model to generate *AtlasifySR+E*'s estimates. In this section, we describe this machine learning approach and evaluate the accuracy of *AtlasifySR+E*'s SR estimates against benchmark SR data sets.

We first ran an experiment to validate the performance of our implementations of SR measures from previous work. Following standard practice, we evaluated each implementation by comparing its SR estimates with datasets of human gold standard estimates using Spearman's r_s and Pearson's r (Table 1). These datasets consist of term pairs and associated SR values, which are averaged across all human annotators of a dataset. We used two long-standing SR datasets, *WordSim353* [8] and *MC30* [21], as well as *TSA287* [30] and *Atlasify240*[4], the SR dataset we developed as part of the experiment described in Section 5.2. The results in Table 1 indicate that our implementations are satisfactory, especially given Wikipedia-based SR measures' tendency to fluctuate in accuracy over time [26, 42].

Our approach to combining the estimates of each constituent SR measure was to use a regression model to predict the human gold standard judgments in *WordSim353*, the most common SR dataset in the literature. We then used this trained model[5] to predict the gold standard judgments in the four SR datasets discussed above. The regression model employed a variety of features, including the SR estimates produced by each constituent measure, along with numerous properties of the Wikipedia article corresponding to each term in a term pair (e.g. article length, link density). Our model also included as a feature the entropy of the word sense disambiguation task required to identify matching articles for each term. *AtlasifySR+E* uses a pairwise maximization approach for

[4] *Atlasify240* is available for download at http://www.cs.northwestern.edu/~ddowney/data_code.html

[5] Although *MC30* and *WordSim353* are frequently simultaneously used to evaluate SR measures, *MC30* is a subset of *WordSim353*. As such, prior to training the model on which we tested on *MC30*, we removed the 30 overlapping pairs from our training set (leaving 323 pairs available for training). We did the same for the 1 overlapping pair with TSA287.

SR Algorithm		MC30		WordSim353		TSA287		Atlasify240	
		r_s	r	r_s	r	r_s	r	r_s	r
WikiRelate	AtlasifySR+E	**.78**	**.82**	.49	.48	.40	.47	.52	.53
	Published	**-**	.57	-	.53	-	-	-	-
MilneWitten	AtlasifySR+E	**.64**	.65	.56	.52	.49	.45	.68	.69
	Published	**.70**	-	.69	-	-	-	-	-
WeightedMW		**.65**	.65	.66	.57	.53	.46	**.74**	**.72**
WAGDirect		**.71**	.73	.64	.58	.49	.53	.60	.56
OutlinkOverlap		**.64**	.67	.52	.42	.48	.42	.61	.51
ESA	AtlasifySR+E	**.74**	.77†	.72	.70†	**.58**	**.62†**	.71	**.72†**
	Published	**.72**	-	.75	-	**-**	-	-	-
TSA (*current SoA*)	Published	-	-	**.80**	-	**.63**	-	-	-
AtlasifySR+E		**.75**	**.81**	**.78‡**	**.76‡**	**.64**	**.68**	**.78**	**.77**
Inter-annotator Agreement		n/a	**.90**	n/a	**.55-.73**	n/a	-	n/a	**.77**

Table 1. The performance of the SR measures considered in this paper, in context with that of their published versions. Where inter-annotator agreement (InterAA) is available, bold indicates results with which we could not detect a significant difference with InterAA using the method in [10] *and* $p < 0.05$. *Where it is not available, bold indicates the top result and those with which we could not detect a significant difference with the top result. InterAA is not included for Spearman's r* (r_s) *due to the prevalence of ties* [43]. *Note that* AtlasifySR+E *is the only measure that is bold in all columns, including those for which there is data for the current state-of-the-art,* TSA. *A ‡ indicates that the model was trained on this dataset. A † indicates that the log of the estimates has been used for improved performance. Finally, a dash means that data was not reported.*

word sense disambiguation [22, 26], wherein word sense candidates are identified using anchor texts.

We found that a boosted implementation of Quinlan's M5 algorithm for smoothed trees of linear regression models achieved good performance using 10-fold cross validation (mean $r = 0.75$ with gold standard values). Among the most predictive features in the model were the SR scores generated by the constituent algorithms and the word sense entropy of the term pair. The constituent SR measure with the most predictive power was *ESA*.

We then evaluated the performance of our new *AtlasifySR+E* measure using the same experimental setup as above. The full results can be seen in Table 1. *AtlasifySR+E* performs better than all Wikipedia-specific measures on every dataset but *MC30* for both correlation metrics, and the *MC30* differences are not significant. Further, we could not detect a statistically significant difference between *AtlasifySR+E*'s Pearson's correlations and the inter-annotator agreement in every case.

We also could not detect a significant difference between the accuracy of SR estimates generated by *AtlasifySR+E* and those generated by *TSA*, which is the current state-of-the-art SR algorithm. *AtlasifySR+E* relies only on Wikipedia data while *TSA* additionally uses exogenous information in the form of a large set of *New York Times* abstracts stretching over decades. This data is language-specific, less accessible than Wikipedia, and less open. *AtlasifySR+E* may thus be preferable to *TSA* in, for example, for-profit settings, non-English contexts, and cross-language information retrieval, a popular application of semantic relatedness (e.g. [29]). Atlasify supports exploratory inquiry in 25 language editions of Wikipedia, and the multilingual nature of *AtlasifySR+E* is a major reason we were able to make Atlasify so universal. We also note that *AtlasifySR+E*'s ensemble approach – improving performance by combining different perspectives on the relatedness between concepts – can incorporate additional perspectives on relatedness, such as *TSA*'s temporal approach and future innovations.

5.2 Explanation Ranking Experiments

Each of *AtlasifySR+E*'s constituent SR+E measures returns a list of explanations ranked by their informativeness (Section 4.2). *AtlasifySR+E* must then consolidate and rank the explanations from each measure into a single list to return to the user when they click on a spatial entity. We approached this explanation ranking task as follows: given a concept pair and the up to six top-ranked (most informative) explanations from the constituent measures, *AtlasifySR+E* is to select the best explanation. *AtlasifySR+E* then iterates, removing the explanation it judged to be most interesting at each iteration and placing it in order in the list of explanations to be returned to the user. In the case of the constituent SR measure whose explanation was placed in the returned list, the next most informative explanation is considered in the subsequent iteration. Solving this ranking problem involved gathering a dataset from human judges and then using this dataset to train, develop, and test a ranker. We describe this effort below.

5.2.1 Data collection

Our training data was based on 268 manually selected concept pairs. Each concept mapped unambiguously to a Wikipedia article, and, following one approach in the literature (e.g. [8, 25]), concept pairs were hypothesized to uniformly cover the spectrum of semantic relatedness. While 28 of these concept pairs come from *WordSim353*, 240 are original pairs not seen before in the SR literature. These 240 pairs make up the *Atlasify240* dataset, which is focused on named entities. Named entities make up a large majority of concepts in spatial reference systems (e.g. "John McCain", "Israel", "Helium"). Existing datasets (e.g. [8, 21, 30]) include relatively few named entities, necessitating new concept pairs for our evaluation.

Each of the most informative (top-ranked) explanations from *AtlasifySR+E*'s constituent SR+E measures was generated for all of the 268 pairs and placed in a Web interface (when there was an explanation available). The interface allowed human annotators to rank the explanations for each pair of articles using drag-and-drop techniques. The presentation order of both the pairs and the explanations were randomized. Prior to ranking explanations for a pair, annotators were required to provide an SR estimate. Following the existing SR literature [25, 30], annotators were able to rank SR on a limited scale, in our case from 0 (not related) to 4 (very related). After ranking the available explanations, annotators were asked if they thought that their top-ranked explanation was a good explanation of a relationship between the two concepts.

Ten annotators finished all pairs. On average, annotators said 66% of their top-ranked explanations were good explanations of the relationship between the two concepts. As hypothesized, *WAGDirect* was by far the best algorithm, with 55% of its explanations being chosen as the best on average. However, *WAGDirect* was only able to produce an explanation in 26.8% of cases because only that many of the article pairs had at least one link between them. *WikiRelate* was lowest performing algorithm, but was still selected 13.8% of time when it was available. The *MilneWitten* algorithms were the most prolific and were each able to generate an explanation for over 80% of the samples.

For 18 (6.7%) pairs, no algorithm was able to generate an explanation. This is to be expected for pairs with very low SR; where there is no relatedness, there is no relationship to explain. Indeed the average mean SR judgment for these pairs was 0.52 (in a 0-4 range). In contrast, the average mean SR judgment for pairs for which all six algorithms generated explanations was 3.78.

5.2.2 Machine Learning

Using the hand-annotated ranks from our data collection process, we developed a dataset that consisted of numerous features for each explanation, including: (1) the SR value estimate from the constituent SR+E measure, (2) the textual quality of the explanation (described in Section 5.3), and (3) an indicator of which SR+E constituent measure produced the explanation. For each pair, we assigned the explanation with the lowest (i.e. best) mean rank a "1" and every other explanation a "2". We trained a ranker to predict the best ("1") explanation using SVMRank [18].

The results of this experiment can be found in Table 2. We report these results in terms of *coverage*, which is the percentage of pairs for which one or more explanations were available, and *precision*, which is the percentage of pairs for which *AtlasifySR+E* correctly identified the best explanation (when one or more were available).

Using 10-fold cross-validation, our best performing model had a precision of 56%, which is significantly better than random guessing ($\chi^2 = 13.2$, $p < .01$) and only 2% lower than mean inter-annotator agreement (58%). In other words, the model predicts the best explanation almost as well as humans agree on the best explanation. The difference between the model and the inter-annotator agreement is in fact not significant ($\chi^2 = 0.51$, $p = .48$). Moreover, this model results in a slightly better precision (insignificantly so) than *WAGDirect*, the best SR+E algorithm for explanations, and has a much higher coverage; it can return an explanation when *any* of the constituent algorithms can find an explanation. In our experiment, this was 93.3% of the time, compared to *WAGDirect*'s 26.8%.

It is important to note that a model based only on which SR+E method was used ("Measure indicators only") performs nearly as well as the full model, and the difference between them is not significant ($\chi^2 = 1.27$, $p = .15$). That is, the relative performance of the constituent SR+E explanation generators accounts for most of the predictive power of our ranking model.

Model features	*Precision*	*Coverage*
All features	56%	93%
Measure indicators only	51%	93%
Random	39%	93%
WAGDirect (Highest Precision Single SR Measure)	55%	27%
MilneWitten (Highest Coverage Single SR Measure)	35%	80%

Table 2. Results of our explanation ranking experiment.

5.3 Quality of Mined Text

The final machine learning experiment we will discuss assesses the quality of text mined from Wikipedia. This quality assessment, along with relationship strength estimates, is used to calculate the informativeness of the explanations for each of *AtlasifySR+E*'s constituent SR+E measures (Section 4). This informativeness is then used to rank explanations within each individual measure.

Hand-annotated data was supplied by four annotators, each of whom rated 500 snippets on a scale from 0 to 4 according to the quality of the natural language. Each snippet describes one "leg" of a relationship (e.g. Figures 10a and 10c have two snippets, while 10b has one). Quality was assessed using several factors, including readability and clarity of relationship described. Inter-annotator reliability was $r = 0.51$ (calculated with Fisher's z-value transformation [41]).

For training, each snippet was assigned two types of features: syntactic (e.g. lack of a verb) and contextual (e.g. at the top of the page). After experimenting with a variety of regression models, we found a linear regression model to be the most accurate. Using 10-fold cross-validation, this model was able to achieve a mean correlation of $r = 0.32$ (Table 3). While the combined model outperforms a model trained on only a single type of feature, models trained on either type of feature alone were not found to have significantly worse predictive power.

Model features	*r* with gold standard
All features	0.32
Contextual features only	0.29
Syntactic features only	0.24
Human Inter-rater Agreement	*0.51*

Table 3. The results of our text snippet quality experiment.

6. CONCLUSION AND FUTURE WORK

In this paper, we showed how thematic cartography can be used for domain-neutral exploratory search, demonstrated this process in a working system called Atlasify, presented two use cases, and evaluated its key functions. We have also made three additional contributions that have implications outside of exploratory search. First, we presented explicit spatialization, which brings the benefits of online mapping and geographic information retrieval into domains beyond geography. Second, we introduced explanatory semantic relatedness (SR+E) measures, which extend popular SR measures to make them user-understandable. We built and evaluated an SR+E measure that produces explanations significantly better than those of baseline approaches. Finally, through its method of merging the benefits of its six constituent SR measures, our new SR+E measure produces SR estimates that are statistically indistinguishable from the state-of-the-art without relying on language specific or proprietary data.

Future work includes applying ES in traditional GIR applications and SR+E in traditional SR applications. We are also incorporating into *AtlasifySR+E* DBpedia's structured relationships, which are concise but of limited coverage. Finally, we are preparing a lab-based evaluation of Atlasify, an essential next step [40], as well as looking to the deploy the system more widely to see how our approach is leveraged by real users.

ACKNOWLEDGEMENTS

The authors thank Jeffrey Geiger for his spatiotagging work, our annotators, and our reviewers. Thanks also to Patti Bao and

Stephanie Hille. This work was supported by an NSF Graduate Fellowship (first author) and NSF Grant IIS-101675.

REFERENCES

[1] Bergstrom, T. and Karahalios, K. 2009. Conversation clusters: grouping conversation topics through human-computer dialog. *CHI '09*.

[2] Bertin, J. and Berg, W.J. 1989. *Semiology of Graphics*. University of Wisconsin Press.

[3] Bozzon, A., Brambilla, M., Ceri, S. and Fraternali, P. 2010. Liquid query: multi-domain exploratory search on the web. *WWW '10*.

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

[5] Card, S.K., Mackinlay, J.D. and Shneiderman, B. 1999. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann.

[6] Dodge, M., Kitchin, R. and Perkins, C. 2011. Introduction to The Map Reader. *The Map Reader: Theories of Mapping Practice and Cartographic Representation*. Wiley.

[7] Eisenstein, J., O'Connor, B., Smith, N.A. and Xing, Eric P. 2010. A Latent Variable Model for Geographic Lexical Variation. *EMNLP '10*.

[8] Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G. and Ruppin, E. 2002. Placing Search in Context: The Concept Revisited. *ACM Transactions on Information Systems*. 20, 1 (2002), 116–131.

[9] Gabrilovich, E. and Markovitch, S. 2007. Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. *IJCAI '07*.

[10] Gabrilovich, E. and Markovitch, S. 2009. Wikipedia-based Semantic Interpretation for Natural Language Processing. *JAIR*. 34, (2009), 443–498.

[11] Goodchild, M.F., Yuan, M. and Cova, T.J. 2007. Towards a general theory of geographic representation in GIS. *IJGIS*. 21, 3 (2007), 239–260.

[12] Google Correlate: *http://correlate.googlelabs.com/*. Accessed: 2011-07-28.

[13] Hearst, M.A. 2009. *Search User Interfaces*. Cambridge University Press.

[14] Hecht, B. and Gergle, D. 2010. The Tower of Babel Meets Web 2.0: User-Generated Content and Its Applications in a Multilingual Context. *CHI '10*.

[15] Hecht, B. and Raubal, M. 2008. GeoSR: Geographically explore semantic relations in world knowledge. *AGILE '08*.

[16] Hecht, B. and Schöning, J. 2008. Mapping the Zeitgeist. *GIScience '08 (Extended Abstracts)*.

[17] Hornbæk, K. and Frøkjær, E. 1999. Do Thematic Maps Improve Information Retrieval? *INTERACT '99* (1999), 1–8.

[18] Joachims, T. 2006. Training Linear SVMs in Linear Time. *KDD '06*.

[19] Jones, C.B., Purves, R.S., Clough, P.D. and Joho, H. 2008. Modelling Vague Places with Knowledge from the Web. *IJGIS*. 22, 10 (2008), 1045–1065.

[20] MacEachren, A.M. 1982. The Role of Complexity and Symbolization Method in Thematic Maps. *Annals of the Assoc. of American Geographers*. 72, 4 (1982), 495–513.

[21] Miller, G.A. and Charles, W.G. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes*. 6, 1 (1991), 1–28.

[22] Milne, D. and Witten, I.H. 2008. An Effective, Low-Cost Measure of Semantic Relatedness Obtained from Wikipedia Links. *WIKI-AI '08*.

[23] Milne, D. and Witten, I.H. 2008. Learning to link with Wikipedia. *CIKM '08*.

[24] Montello, D.R. 2002. Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches. *CAGIS*. 29, 3 (2002), 283–304.

[25] Pedersen, T., Pakhomov, S.V.S., Patwardhan, S. and Chute, C.G. 2006. Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*. 40, 3 (2006), 288–299.

[26] Ponzetto, S.P. and Strube, M. 2007. Knowledge Derived From Wikipedia For Computing Semantic Relatedness. *JAIR*. 30, (2007), 181–212.

[27] Popescu, A. and Grefenstette, G. 2010. Mining User Home Location and Gender from Flickr Tags. *ICWSM '10*.

[28] Popescu, A. and Grefenstette, G. 2010. Spatiotemporal mapping of Wikipedia concepts. *JCDL '10* (2010), 129–138.

[29] Potthast, M., Stein, B. and Anderka, M. 2008. A Wikipedia-based multilingual retrieval model. *ECIR '08*.

[30] Radinsky, K., Agichtein, E., Gabrilovich, E. and Markovitch, S. 2011. A Word at a Time: Computing Word Relatedness using Temporal Semantic Analysis. *WWW '11*.

[31] Rajaraman, A. 2009. Kosmix: High-Performance Topic Exploration using the Deep Web. *Proceedings of the VLDB Endowment*. 2, 2009 (2009), 1524–1529.

[32] Risch, J.S., Rex, D.B., Dowson, S.T., Walters, T.B., May, R.A. and Moon, B.D. 1997. The STARLIGHT information visualization system. *INFOVIS '97*.

[33] Rose, D.E. and Levinson, D. 2004. Understanding user goals in web search. *WWW '04*.

[34] Sheesley, B. 2009. Data Probing and Info Window Design on Web-based Maps. *Axis Maps Blog*. http://www.axismaps.com/blog/2009/07/data-probing-and-info-window-design-on-web-based-maps/. Accessed: 2012-06-04.

[35] Shneiderman, B. 1996. The eyes have it: a task by data type taxonomy for information visualizations. *IEEE Symposium on Visual Languages '96* (Sep. 1996), 336 –343.

[36] Skupin, A. and Fabrikant, S.I. 2003. Spatialization Methods: A Cartographic Research Agenda for Non-geographic Information Visualization. *CAGIS*. 30, 2 (2003), 95–115.

[37] Slocum, T.A., McMaster, R.B., Kessler, F.C. and Howard, H.H. 2009. *Thematic Cartography and Geovisualization*. Prentice Hall.

[38] Strube, M. and Ponzetto, S.P. 2006. WikiRelate! Computing Semantic Relatedness Using Wikipedia. *AAAI '06*.

[39] White, R., Muresan, G. and Marchionini, G. 2006. Evaluating Exploratory Search Systems. *SIGIR '06 Workshop on Evaluating Exploratory Search* (2006).

[40] White, R., Roth, R. and Marchionini, G. 2009. *Exploratory search: beyond the query-response paradigm*. Morgan & Claypool.

[41] Zesch, T. and Gurevych, I. 2006. Automatically creating datasets for measures of semantic relatedness. *ACL-Workshop on Linguistic Distances*.

[42] Zesch, T. and Gurevych, I. 2010. The More the Better? Assessing the Influence of Wikipedia's Growth on Semantic Relatedness Measures. *LREC '10.*

[43] Zesch, T. and Gurevych, I. 2009. Wisdom of crowds versus wisdom of linguists – measuring the semantic relatedness of words. *Natural Language Engineering*. 16, 1 (2009), 25–59.

[44] Zesch, T., Gurevych, I. and Mühlhäuser, M. 2007. Analyzing and accessing Wikipedia as a lexical semantic resource. *Data Structures for Linguistic Resources and Applications*.