UC Santa Barbara
UC Santa Barbara Previously Published Works

Title

Special issue introduction: Spatial approaches to information search

Authors

Ballatore, Andrea

Kuhn, Werner

Hegarty, Mary

et al.

Peer reviewed

Title: Spatial approaches to information search

Authors: Andrea Ballatore (Birkbeck, University of London and University of California, Santa Barbara), Werner Kuhn (University of California, Santa Barbara), Mary Hegarty (University of California, Santa Barbara), Ed Parsons (Google)

Abstract

Searching for information is a ubiquitous activity, performed in a variety of contexts and supported by rapidly evolving technologies. As a process, information search often has a spatial aspect: spatial metaphors help users refer to abstract contents, and geo-referenced information grounds entities in physical space. While information search is a major research topic in computer science, GIScience and cognitive psychology, this intrinsic spatiality has not received enough attention. This article reviews research opportunities at the crossroad of three research strands, which are (1) computational, (2) geospatial, and (3) cognitive. The articles in this special issue focus on interface design for spatio-temporal information, on the search for qualitative spatial configurations, and on a big-data analysis of the spatial relation "near".

Keywords: information search; spatial information; cognitive search; geographic information retrieval

Introduction

Information search is a major component in many human activities. Web search engines process billions of queries every day and determine the visibility and accessibility of much online content. Scientists search for meaningful patterns in increasingly large datasets, while consumers search for products and services among many available options. The search for information has been tightly intertwined with a spatial dimension (Todd et al. 2012). Human and artificial agents traverse heterogeneous information spaces searching for entities and their relations, in an analogy with how biological organisms explore their physical environment to search for sources of nourishment.

In this sense, there is a spatial component at the core of information search. Most search technologies rely on spatial metaphors: for instance, we refer to *going to* websites to search for fragments in an overwhelmingly large abstract space of messages, documents, images, and videos. Computing technologies spatialize abstract pieces of information into tangible interfaces and layouts. The physical geographic space grounds information and helps refine search strategies, relying on the location of entities on the Earth's surface to assess their relevance.

While this spatial dimension of information search is pervasive in many disciplines, including computer science, geographic information science (GIScience), and cognitive psychology, there has been limited interaction and cross-fertilization of these fields. Hence, this special issue explores precisely the spatial dimensions of, and approaches to information search from several

interdisciplinary perspectives. To deal with this broad area of inquiry, we focus on the interplay between computational, geospatial, and cognitive research strands (Ballatore et al. 2015). These strands are thoroughly interconnected, and we do not propose them as a clear thematic partition, but rather as centres with porous peripheries.

By the computational strand (1), we refer to approaches to information search that are grounded in mathematical formalization and algorithms (see Section 2). Starting from seminal work in artificial intelligence (Russell et al. 2009), the computational strand focuses on developing efficient methods to explore large information spaces, when exploring all possibilities is not feasible. Computational approaches have radically transformed information search, resulting in the engineering of database management systems and search engines, in the rich area of information retrieval. More recently, the explosion of "big data" has opened up novel informational spaces, characterised by heterogeneity and varying levels of semantic structure.

The geospatial strand (2) hinges on a particular search space, i.e. geographic space, intended as the space near the surface of the Earth. This space is particularly important as it provides a unified ground to anchor disparate pieces of information, enabling search for information related to human and non-human phenomena occurring in space and time, from the local to global scale. Both computer science and GIScience have engaged with search techniques tailored to geographic dimensions of information (Murdock et al. 2014; Jones & Purves 2008). In particular, the area of geographic information retrieval (GIR) has tackled computational challenges such as the determination of geographic relevance of text documents, and the disambiguation of place names (Section 3).

Finally, the cognitive strand (3) takes a different tack on information search, focusing on how the human cognitive apparatus searches for information both in physical space, for example using mechanisms of visual search (Eckstein 2011), and how it retrieves information from memory (Todd et al. 2012). Knowing how humans perform information searches is arguably crucial to design better information retrieval systems, to support interaction design and geovisual analytics, and to spatialize abstract spaces effectively (Pirolli, 2007). A complementary issue concerns the impact of increasingly pervasive search technologies on how humans cognize the physical-geographic reality (Section 4).

Without pretension to exhaustiveness, the remainder of this article identifies themes and threads at the intersection of these broad disciplinary areas, highlighting linkages, synergies, fractures, and, above all, promising research gaps that appear ripe for an interdisciplinary agenda. As observed in a specialist meeting held in Santa Barbara in December 2014 (Ballatore et al. 2015), these complementary perspectives can interact more extensively to reap mutual benefits for scientific and technological advances.

The spatial dimension of information search

In computer science and in artificial intelligence, search is seen as foundational for problem solving and planning (Russell et al., 2009). Problems are conceived as abstract spaces, in

which intelligent agents search for solutions, without having the possibility of an exhaustive search, having to rely on heuristics to narrow down the search to a manageable size. Papadimitriou (2014) aptly noted that many fundamental problems in computer science are search problems, defined as "given an input, call it *x*, find a solution *y* such that *x* and *y* stand in a particular relation to each other that is easy to check" (p. 15881). Many of such fundamental search problems consist of finding paths in spaces structured as networks (e.g., shortest path, Hamilton path, Clique, and the Min-cut). In this context, information search is seen as a complex process reducible to a sequence of basic computational operations (e.g., read/write), identifiable through an algorithm (or, in difficult classes of problems, not decidable).

As Franklin and Andrade point out (Franklin and Andrade, in Ballatore et al. 2014, pp. 16–18), the remarkable increase of available memory in computational systems is also changing which data structures are appropriate to solve search problems. While many retrieval operations have become trivial, finding objects that co-occur in the search space is still a challenge in very large databases. Research in database management systems has generated data structures and indices to enable more efficient search, for spatial and non-spatial dimensions, such as array and non-relational databases, going beyond relational databases that have dominated the landscape for 40 years (Brown, 2010). Similarly, linked data and semantic web technologies offer a platform to integrate disparate and heterogeneous data spaces into a unified, searchable space, structured as a dynamic network of triples (Kuhn et al., 2014).

Geographic information search

Geographic space is particularly important in information search, as it pervades informational content, providing ground for linking different data spaces. Among all search spaces, geography emerges as a particularly important one. Core concepts of spatial information, such as objects, fields, events, and networks provide a suitable conceptual infrastructure to organize, integrate, and search geographic information (Kuhn, 2012). Geospatial information has also been proposed as a facilitator for discovery and interdisciplinary collaboration in the context of scientific libraries (Lafia et al, 2016).

In GIScience, three dimensions of information (spatial, temporal, and thematic) remain ubiquitous in framing the complexities of geographic information, as well as search of geographic information (Yuan, 1999). However, in the social sciences and the digital humanities, a new focus has emerged on the notion of *place*, i.e. a socially and culturally constructed object, rather than a merely topological and spatial entity associated with some thematic description. For example, the description of a city as a place includes a nexus of complex human agents, activities, processes, and relations, well beyond the enumeration of the location of its roads and buildings. The advantages of indexing information with respect to place is apparent for exploratory search as well as for analysis (Grossner, in Ballatore et al. 2014, pp. 26–28).

As information is increasingly consumed through mobile devices, the geo-location of the users has gained prominence to refine the search process, as well as an important element of user-

generated content (Graham et al, 2014). In recent years, novel sources of geographic information have erupted, resulting in large and dynamic datasets of geo-tagged photographs, messages, videos, and check-ins (Murdock, 2014). To extract insights from such information and make the information more searchable, computational models are including explicit locational information, increasing relevance and level of personalization.

One of the most prominent efforts to support geographic information search in GIScience can be seen in geographic information retrieval (GIR) (Jones & Purves, 2008). This interdisciplinary area focuses on the geographic content of text documents, harnessing concepts and techniques from computational linguistics and natural language processing. The cogency of GIR lies also in the increased availability of very large text corpora that contain rich spatio-temporal information (Michel et al., 2011). A notable challenge in GIR is the recognition and disambiguation of place names in text, which remains difficult for fully automated systems. Moreover, as pointed out by Purves (Purves, in Ballatore et al. 2014, pp. 72–76), user interfaces for geographic search have not improved substantially beyond the display of results as points (or polygons) on generic base maps. Although difficult to obtain, query logs in search engines are still an unsurpassed tool to better understand how users interact and express spatial needs on real systems. Given the still limited interdisciplinary interaction between GIR and spatial cognition, tangible benefits could be brought about to better understand how users formulate spatial queries and how they acquire spatial knowledge.

Information search and spatial cognition

As search is a fundamental activity of human and animal mental life, cognitive psychologists have investigated the structures and processes that govern it. As Todd et al. (2012) noted in their comprehensive survey, organisms perform similar searches in a variety of contexts, highlighting the commonalities (and indeed differences) between searching in visual, aural, spatial, social, and memory spaces. The theory of information foraging draws a strong analogy between search for food in the physical environment and search for information in abstract and digitally-mediated spaces, based on the evolutionary assumption that search strategies evolved first to ensure successful physical foraging – and therefore survival (Pirolli, 2007). Spatialization is thus an important methodology to make abstract spaces cognizable and searchable in an intuitive way.

In a societal context, where search in digital informational spaces has become crucial to carry out daily tasks, understanding how information search occurs at a deep, cognitive and neural level can provide insights to build more effective search tools. While human-computer interaction and cognitive psychology have a long and fruitful history (Card et al, 1984), an area where more interplay between the three strands is needed is geovisual analytics, where visual search (Eckstein, 2011), and spatial language (Matlock et al, 2014) are paramount. In summary, little interaction has occurred between the cognitive and other research strands to systematically study and exploit the spatial dimensions of information search as a cognitive task.

Challenges and opportunities

The interdisciplinary discussions at the Specialist Meeting in Santa Barbara (Ballatore et al, 2015) have identified a number of promising research themes and questions on information search at the intersection of the computational, geospatial and cognitive strands. Hoping to stimulate further interest beyond this special issue, we summarize them here.

Spaces and places. The humanistic notion of place is multi-faceted and complex, and yet we cannot easily search for places beyond very few and simplistic thematic dimensions (e.g., "cities with more than a million inhabitants"). Better "platial" models are needed to include the notion of place into geographic information systems, which are traditionally (and successfully) built on topological spaces. The challenges to place computing include the ad hoc, subjective, and mutable nature of place. To a large extent, the information retrieval community still ignores space and place, and more efforts from GIScience are needed to make these perspectives more central to research on information search. In particular, articulating and working on specific problems of place-based search appears to be an opportunity for collaboration.

Visualization of big spatial data. To provide better organization of knowledge beyond lists of ranked documents and traditional pins-on-maps visualizations, new visualization methods are needed. From a cognitive perspective, knowledge about mental representations of geographic and abstract spaces is essential to devise more effective approaches to exploring, summarizing, and uncovering meaningful patterns in large datasets. This challenge can benefit from developments in database technology, such as non-relational, column, and array database management systems in addition to research on how humans represent and search both physical and information spaces.

Models of human search behavior. More research in cognitive psychology is needed to further illuminate the strategies and heuristics deployed in search behavior in physical and information spaces, which would deepen our understanding of how humans search for patterns in stimuli and in memory. This information in turn could be used to develop information systems that build on and augment human search abilities.

Benchmarking exploratory search. Compared with task-oriented search, the evaluation of exploratory search is more challenging, because it is difficult to establish objective criteria of success. It would be valuable to design and curate test collections to be used across different research communities. To date, there is a lack of benchmark collections that allow evaluations, hindering reproducibility and comparison of methods to explore informational spaces. The visual dimension, for example through the collection of eye movements, can be used to evaluate users' search strategies and behavioral patterns.

Georeferencing quality. While commercial and open-source tools for georeferencing are available, their quality varies dramatically. Better benchmarking and evaluations are needed to support search for geographic information effectively. Mainstream search engines need better topological and geographic knowledge bases to produce more meaningful results. For example,

a Google search for "distance between Italy and France" returns 1,298 km, ignoring the topological structures of the two adjacent countries, using their arbitrary centroids. In this sense, deciding when a point location is adequate to solve a problem and when extended footprints are needed is a largely unsolved problem.

Vagueness and ambiguity in spatial hierarchies and relations. Geospatial search involves the use of spatial terms, which are often intrinsically vague and context-dependent. Notably, the definition of nearness varies depending on the context, and place name disambiguation is a hard problem, especially for vernacular place names not encoded in a gazetteer. As search in the geographic domain is strongly affected by scale, organizing content in hierarchies is beneficial. However, spatial and thematic hierarchies constitute a challenge for evaluation. These hierarchies should be made more explicit for the user, in order to collect relevance feedback. Similarly, the development of multi-scale, context-sensitive spatial relations has the potential for greatly improving search approaches.

Search in spatio-temporal networks. Many human and natural systems, such as urban transit and social media, can be conceived as networks whose spatial structure changes over time. Their properties are emerging from interdisciplinary research and novel techniques are needed to search efficiently for paths, events, patterns, clusters, and outliers in these complex networks. They will bridge established strands of network analysis, such as social network analysis, with spatial and time series analysis.

Effects of search technologies on spatial cognition. The pervasive availability of search technology is redefining the process of retrieval of geographic information, limiting the need for memorization. Beside anecdotal evidence, little is known about how this new technological landscape impacts spatial cognition. Fruitful investigations might focus on psychological aspects, such as spatial awareness and wayfinding abilities, as well as on more social, cultural, and political dimensions of how the geographic world is collectively imagined and accessed.

Unstructured and subjective spaces. Current spatial search is largely confined to structured spatio-temporal data, and ideally search should be possible across large volumes of unstructured spatial data, gathered from social media and other web sources (Hoffart et al. 2013). Thanks to recent advances in natural language processing and machine learning, subjective experiences, emotions, and opinions can become novel search spaces, unlocking new understandings of social and urban dynamics.

Reference systems for abstract spaces. Web maps and time sliders provide a widely used mechanism to consume information structured in the geographic space, but what about abstract spaces, such as conceptual spaces (Gärdenfors 2004)? We need more explicit semantic reference systems for better ontological organization of search spaces. In this context, the metaphor of the map projection can be deployed to represent multiple spatial representations of the same abstract spaces, guiding the development of coordinates systems, and the assessment of distortions in these culturally embedded informational spaces. Cognitive

research on how people conceptualize information spaces may also lead to the development of other usable technologies.

Type instantiation. In geographic information retrieval (as in other searches), queries often refer to instances of geographic entities by referring to their type (e.g., "the beach next to University of California, Santa Barbara" when referring to Goleta Beach). Spatial reasoning and geographic knowledge are needed to resolve this type of indirect referencing, expanding traditional techniques of co-reference resolution.

Search for aggregates and similarities. Searching for individual database records matching a set of criteria is not a notable challenge anymore, even in very large datasets. However, the search for complex aggregates, such as the co-occurrence of events in space and time is still challenging, particularly when facing very large and diverse data sources. Such aggregates include city neighborhoods, large public events, and trajectories. Spatio-temporal datasets can also be conceptualized as special kinds of aggregates, stored in data catalogues. In an ecological approach to information search, the space to be searched is that of multiple interactions between entities, stressing the need to be able to express and solve complex queries for spatial, temporal, and thematic aggregates that emerge in physical and abstract spaces alike. Searching for similar aggregates also represents a worthwhile challenge, as aggregates rarely present exact structures and need fuzzier mechanisms for comparison.

Summary of the special issue

The articles included in this special issue provide stimulating perspectives centered on the spatial dimensions and approaches of information search, addressing some of the aforementioned challenges. Bruggmann and Fabrikant (2016**)** explore the potential of spatial information search for the digital humanities. Harnessing techniques from GIScience, GIR, and the more recent area of geovisual analytics, they propose an interdisciplinary methodology to design usable interfaces for spatio-temporal analysis. In their case study, a text corpus containing articles about Swiss history is processed with computational tools. The resulting spatio-temporal references provide the data to be consumed in an interactive interface, using spatialization to represent thematic information, and providing detailed guidelines to design spatial interfaces, aimed at the consumption and exploration of geographical and historical datasets. The method successfully identifies improvements for the complex interface design, aimed at the reduction of the adoption barrier of spatio-temporal search systems.

From the perspective of qualitative spatial reasoning, Fogliaroni et al. (2016) focus on spatial configuration search, an area largely neglected by current geographic information systems. When a user wants to identify configurations of objects in qualitative terms, their system solves the query by formalizing it as a set of qualitative spatial predicates of arbitrary size. These spatial constraints are then propagated through a hypergraph containing the dataset expressed as qualitative predicates, identifying suitable solutions that capture potentially very complex aggregates of spatial entities holding specific relations.

Information needs are often expressed in ambiguous and vague natural language (e.g., "hotels near the city center"). In their article, Derungs and Purves (2016) take a big data approach to study these vague spatial relations in informational and geographical spaces. Noting that the interpretation of spatial queries in natural language remains a challenge, they inspect a large dataset of linguistic sequences based on billions of web pages (the Microsoft Web N-grams) focusing on the spatial relation "near". Their work provides a method to extract knowledge from n-grams, which are potentially powerful resources, but difficult to disambiguate to reduce noise and misinterpretation of linguistic tokens. This investigation provides new empirical evidence for the asymmetry and other characteristics of this ubiquitous spatial relation, demonstrating the potential of more interaction between computational, geospatial and cognitive research on information search.

References

Bruggmann, A & Fabrikant, S. (2016): How does GIScience support spatio-temporal information search in the humanities?, Spatial Cognition & Computation.

Ballatore, A., Hegarty, M., Kuhn, W., & Parsons, E. (2015). Spatial Search, Final Report. Santa Barbara, CA. Retrieved from https://escholarship.org/uc/item/33t8h2nw.

Ballatore, A., Kuhn, W., Hegarty, M., & Parsons, E. (Eds.). (2014). Position Papers, 2014 Specialist Meeting — Spatial Search. Santa Barbara, CA. Retrieved from http://escholarship.org/uc/item/0h014085

Brown, P. G. (2010). Overview of SciDB: Large scale array storage, processing and analysis. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (pp. 963-968). New York: ACM.

Card, S. K., Newell, A., & Moran, T. P. (1983). The psychology of human-computer interaction. Hillsdale, NJ: L. Erlbaum Associates

Derungs, C. and Purves, R.S. (2016) Mining Nearness Relations from a N-Grams Web Corpus in Geographical Space, Spatial Cognition & Computation.

Eckstein, M. P. (2011). Visual search: A retrospective. Journal of Vision 11(5):14, 1–36.

Gärdenfors, P. (2004). Conceptual spaces: The geometry of thought. Cambridge, MA: MIT Press.

Graham, M., Schroeder, R., & Taylor, G. (2014). Re: Search. New Media & Society 16(2): 187–194.

Hoffart, J., Suchanek, F. M., Berberich, K., & Weikum, G. (2013). YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Artificial Intelligence, 194, 28-61.

Jones, C. B., & Purves, R. S. (2008). Geographical information retrieval. International Journal of Geographical Information Science 22(3): 219–228.

Kuhn, W. (2012). Core concepts of spatial information for transdisciplinary research. International Journal of Geographical Information Science, 26(12), 2267-2276.

Kuhn, W., Kauppinen, T., & Janowicz, K. (2014). Linked Data - A Paradigm Shift for Geographic Information Science. In M. Duckham, E. Pebesma, K. Stewart, & A. U. Frank (Eds.), Geographic Information Science. (pp. 173–186). Berlin: Springer.

Lafia, S., Jablonski, J., Kuhn, W., Cooley, S., & Medrano, F. A. (2016). Spatial discovery and the research library. Transactions in GIS, 20(3), 399-412.

Matlock, T., Castro, S. C., Fleming, M., Gann, T. M., & Maglio, P. P. (2014). Spatial Metaphors of Web Use. Spatial Cognition & Computation 14(4): 306–320.

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., et al. (2011). Quantitative analysis of culture using millions of digitized books. Science, 331(6014), 176–182.

Murdock, V. (2014). Dynamic location models. In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (pp. 1231–1234). ACM.

Fogliaroni, P, Weiser, P., Hobel, H. (2016) Qualitative Spatial Configuration Search, Spatial Cognition & Computation.

Papadimitriou, C. (2014). Algorithms, complexity, and the sciences. Proceedings of the National Academy of Sciences. http://doi.org/10.1073/pnas.1416954111

Pirolli, P. (2007). Information foraging theory: Adaptive interaction with information. Oxford, UK: Oxford University Press.

Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., & Edwards, D. D. (2009). Artificial intelligence: A modern approach (3rd edition). New York: Pearson.

Todd, P. M., Hills, T. T., & Robbins, T. W. (Eds.). (2012). Cognitive Search: Evolution, Algorithms, and the Brain. Cambridge, MA: MIT Press.

Yuan, M. (1999). Use of a Three- Domain Representation to Enhance GIS Support for Complex Spatiotemporal Queries. Transactions in GIS, 3(2), 137-159.