Schlieder, Christoph; Henrich, Andreas

Spatial grounding with vague place models

Date of secondary publication: 25.02.2025

Accepted Manuscript (Postprint), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-1066614

Primary publication

Schlieder, Christoph; Henrich, Andreas (2011): Spatial grounding with vague place models, in: SIGSPATIAL special, New York, NY: ACM, Vol. 3, Nr. 2: Special Issue on Geographic Information Retrieval, pp. 20–23, doi: 10.1145/2047296.2047301.

Spatial Grounding with Vague Place Models

Christoph Schlieder and Andreas Henrich

Faculty of Information Systems and Applied Computer Sciences,
University of Bamberg, Germany
{christoph.schlieder, andreas.henrich}@uni-bamberg.de

Spatial descriptors and place models

Vagueness arises as a problem in the context of grounding named entities, a central task in geographic information retrieval (GIR) which consists in establishing the entities' "denotation with respect to the world or a model" [8]. We use the term *spatial grounding* in a slightly more general sense to refer to the association of any type of *spatial descriptor* with a *place model*. In the simplest case, the descriptor consists of a toponym and the place model is provided by a (vague) geographic region. More complex descriptors make use of prepositional phrases or even a text span comprising many sentences, e.g. a travel blog entry. However, descriptors are in no way limited to natural language expressions. A frequently used descriptor in mobile GIR is the geographic position of the user as part of the specification of an information need. Furthermore, spatial activities such as taking photographs can also act as descriptors for places models ("tell me more about the focused object"). Different place models for representing the vagueness of regions have been proposed including supervaluation semantics and qualitatively augmented fuzzy footprints [12]. In this paper we concentrate on approaches to grounding that exploit diverse information sources, for instance, topographic data or the spatial behavior of a user community, to determine place models which are sometimes more complex than point sets.

Regions with vague boundaries

Most precisely delimited geographic entities are created by institutions such as a national cadastre or mapping agency which have a mandate to define spatial boundaries and to control the naming process. Despite the activities of the United Nations Expert Group on Geographic Names, not all countries have established national geographical names authorities. Where such authorities exist, it is far from clear whether the citizens are acquainted with the official meaning of all toponyms [5]. Furthermore, many place concepts used in everyday communication are not controlled by a naming authority. A Web search for holiday resorts at the "French Riviera" could refer to a region which extends westwards to St. Tropez or Hyères or even Cassis and include just the seaside or also different parts of the inland. Geographic regions for which no single precise boundary exists are considered vague regions in the literature on GIR (e.g. [9]). This understanding calls for some clarification.

Firstly, a boundary is always defined at a certain spatial scale. Strictly speaking there are no precise geographic boundaries because even the delineation of a parcel of land by a surveyor is exact only within the limits of positional accuracy and precision. Frank [1] shows that scale combines information about the error and the finiteness of measurements into a single characteristic which provides an approximation of the size of the smallest localizable regions (e.g. $50m^2$ for 1:50,000). We hold that the basic GIR task, the mapping of spatial information needs onto spatial information sources, does not involve regions which are vague per se. Instead, vagueness is task dependent and arises from a mismatch of scales between information needs and information sources where the smallest units defined at the *spatial task scale* are not localizable with sufficient precision in the place model. A place model of the French Riviera which is uncertain about whether or not to include Cassis is still a precise model for the task of searching information about conversion rates because it permits determining that the local currency is the Euro. The same model is to be considered vague, however, with respect to searches for holiday resorts. Note that depending on the task, it might be necessary to replace geodetic distance by distance measurements in a network or a partonomy.

Secondly, vagueness and uncertainty are often confounded. We may know that a region such as a piece of real estate is precisely delimited at a given task scale without being able to trace its boundary because of insufficient sampling, disagreement amongst experts or for other reasons. This uncertainty is often interpreted as a kind of vagueness. Montello et al. [10] introduce the term epistemological vagueness to distinguish this case from ontological vagueness which applies when the membership to a category comes in degrees. Whether a part of a rocky surface is categorized as "granite", "granodiorite" or "diorite" depends on its composition, on the percentage of feldspar and silica it contains. Note that two membership functions encoding uncertainty and vagueness are needed to distinguish between the cases with high positional uncertainty and those with low categorization confidence.

Thirdly, boundaries arise from physical discontinuities (e.g. water bodies) or social conventions (e.g. real estate) or a mixture of both. In ontological modeling, the former type is referred to as *bona fide* boundary, the latter as *fiat* boundary. It has been argued that a different type of topological reasoning is needed for fiat boundaries since they do not support the open/closed distinction of classical point set topology [13]. Fiat boundaries arising from social conventions generally have a rather limited lifetime and often come with a complex history of precursor entities. Many are known only to specific communities. A company which introduces "Denver" as descriptor for a marketing region defined in terms of ZIP code areas creates an ad hoc place name whose denotation may change with the next business year. Thus, two important parameters of place models for regions defined by social conventions are their lifetime and their user community [15].

Physical and Social Grounding

The distinction between bona fide and fiat boundaries reappears in two strands of research on grounding the semantics of spatial regions: physical grounding and social (or behavioral) grounding. Physical grounding exploits the fact that some place models are strongly constrained by the geographic environment. Straumann and Purves [14] found good agreement between their algorithm for delineating a valley floor in a digital elevation model and the vague region associated with the vernacular place name for that valley which they reconstructed from mining touristic Web sites. Their data show that physical grounding can be used as a complementary source of information to improve the models for vague regions obtained from Web mining. This is especially true for places like the Rhine Valley or the Tibetan Plateau whose vernacular meaning is closely linked to geomorphologic features. Geomorphology is not the only kind of information source which can be exploited for physical grounding. Hydrographic data ("Mediterranean Sea") and land cover data ("Black Forest") could provide valuable cues as well.

Social grounding exploits behavioral constraints that are observed within specific user communities. This is of particular interest when moving from traditional document-centered GIR tasks to service-centered tasks like recommending. Schlieder and Matyas [11] show photographs published in Web collections can be used to measure the social visibility of a place and to customize a geographic recommender. The place model consists of the co-selection matrix for points of interest (POI) which records which POI are photographed by the same people. Such a model can be used to generate density surfaces that permit to decide whether a place belongs to the vague touristic conceptualization of, for instance, Amsterdam. However, co-selection data supports more complex tasks than deciding fuzzy membership since it provides information about correlations of membership decisions on the semantic fringe of a place concept which generally depend on the subcommunity a user belongs to. Such a place model supports predictions of the type "whoever gives point *A* a high fuzzy membership value is likely to do the same for point *B"*. It turns out that agreement on rarely photographed POI is a better predictor than agreement on frequently photographed ones.

Spatial tracks recorded by GPS smartphones or recovered from photo time stamps are another type of behavioral data that has been used to build place models. Girardin et al. [3] describe a model of Rome as tourist destination which encodes how visitors move between different POIs. The place model is given by a transition graph which specifies the probabilities for choosing the next POI. PlaceRank, a spatial version of eigenvector centrality, is used to characterize the POIs that are socially most visible [4]. It is characteristic of this type of place model to include information about the temporal sequence of spatial decisions. Both, the co-selection matrix and the transition matrix place model aim at making explicit the network structure of places which is induced by the spatial behavior of user communities. Exploring the relationship between the region-based and the network-based vague place models is a task for future research.

A Research Agenda: The Physical and the Social Visibility of Places

Physical grounding addresses visible features of the geographic environment that constrain the conceptualization of places while social grounding exploits social visibility, that is, the fact that in social processes some actors, topics or places are more salient than others. GIR services need to handle both physical and social visibility, since their users navigate in information spaces and, at the same time, move in geographic environments. Both types of navigation follow a similar type of information economy in which decisions by individuals or groups require highly selective filtering mechanisms to avoid information overload. It is in this context that the meaning of place concepts needs to be understood. A number of research challenges emerge:

- *From vague regions to vague places*: Spatial conceptualizations are tied to user communities and place models should reflect this dependence. The classical membership problem (does point *X* belong to the vague region *R*?) leads to more complex problems: does *A* believe *X* belongs to *R*? Do *A* and *B* share similar beliefs about *X* belonging to *R*? Place models that permit to answer these questions are not necessarily region-based. Different types of network-based models are another promising candidate for describing vague places.
- *From documents to services*: As services become more of a focus in GIR, grounding will need to move beyond named entity recognition in documents to include geographic behaviour as a descriptor for information needs and to develop computational methods for activity recognition. An open issue is how to include activity patterns in the place models.
- *From models to usage*: In IR the quality and complexity of vague place models is not an end in itself. Lastly, the question is whether the representation and understanding of vague places allows us to increase the effectiveness of GIR systems. In [7] we have shown that traditional quality measures for the accuracy of vague place representations

are not directly correlated to their appropriateness for IR purposes. Hence, one important research issue consists of exploiting rich and expressive models for vague places in the retrieval process. Besides a better understanding of geographic information needs − see e.g. [2] or [6] − a probabilistic foundation might be beneficial in this respect.

References

[1] Frank, A. (2009): Why is scale an effective descriptor for data quality? The physical and ontological rationale for imprecision and level of detail. In: G. Navratil (ed.): *Research Trends in Geographic Information Science*. Berlin: Springer, p. 39–62.

[2] Gravano, L.; Hatzivassiloglou, V.; Lichtenstein, R. (2003): Categorizing web queries according to geographical locality. In: D. Kraft (ed.): *CIKM-03 Proceedings of the 12th Int. Conference on Information and Knowledge Management*. New York: ACM Press.

[3] Girardin, F.; Blatt, J.; Calabrese, F.; Dal Fiore, F.; Ratti, C. (2008): Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive Computing* 7 (4), p. 36–43.

[4] Girardin, F.; Vaccari, A.; Gerber, A.; Biderman, A.; Ratti, C. (2009): Quantifying urban attractiveness from the distribution and density of digital footprints. *International Journal of Spatial Data Infrastructure Research* 4, p. 175-200

[5] Helleland, B. (2006): The social and cultural values of geographic names. In: *Manual for the National Standardization of Geographical Names*. New York: United Nations Publication, S. 121–128. http://unstats.un.org/unsd/publication/seriesm/seriesm_88e.pdf

[6] Henrich, A.; Lüdecke, V. (2007): Characteristics of geographic information needs. In: R. Purves et al. (eds.): *GIR-07 Proceedings of the 4th ACM Workshop On Geographic Information Retrieval*, ACM Press, p. 1-6

[7] Henrich, A.; Lüdecke, V. (2009): Measuring similarity of geographic regions for geographic information retrieval. In: Boughanem, M. et al. (eds.): *ECIR-09 Proceedings 31th European Conference on IR Research*, Berlin: Springer (LNCS 5478), p. 781-785

[8] Leidner, J.; Sinclair, G.; Webber, B. (2003): Grounding spatial named entities for information extraction and question answering. In: A. et al Kornai (eds.): *GEOREF '03 Proceedings Workshop on Analysis of Geographic References*, ACL, p. 31–38.

[9] Jones, C.; Purves, R.; Clough, P.; Joho, H. (2008): Modelling vague places with knowledge from the Web. *International Journal of Geographical Information Science* 22 (10), p. 1045-1065.

[10] Montello, Daniel; Goodchild, Michael; Gottsegen, Jonathon; Fohl, Peter (2003): Where's downtown? Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation* 3 (2&3), S. 185–204.

[11] Schlieder C., Matyas C. (2009). Photographing a city: An analysis of place concepts based on spatial choices. *Spatial Cognition and Computation* 9 (3), p. 212-228

[12] Schockaert, S. (2011). Vague regions in geographic information retrieval, *SIGSPATIAL special issue on Geographic Information Retrieval*, this issue

[13] Smith, B.; Varzi, A. (2000): Fiat and bona fide boundaries. *Philosophy and Phenomenological Research* 60 (2), p. 401–420.

[14] Straumann, R; Purves, R. (2008): Delineation of valleys and valley floors. In: T. Cova et al . (eds.): *GIScience-08 Proceedings of the 5th Int. Conference on Geographic Information Science*. Berlin: Springer (LNCS, 5266), p. 320–336.

[15] Vögele, Thomas (2004): *Spatial Information Retrieval with Place Names*. Doctoral dissertation, University of Bremen. elib.suub.uni-bremen.de/ publications/dissertations/E-Diss978_voegele.pdf.