Semantic place localization from narratives*

Simon Scheider
Institute for Geoinformatics, University of Münster
Heisenbergstr. 2
48149 Münster, Germany
simonscheider@web.de

Ross Purves
Department of Geography, University of Zurich - Irchel
Winterthurerstrasse 190
8057 Zürich, Switzerland
ross.purves@geo.uzh.ch

ABSTRACT

Place narratives provide a rich resource of learning how humans localize places. Place localization can be done in various ways, relative to other spatial referents, and relative to agents and their activities in which these referents may be involved. How can we describe places based on their spatial and semantic relationships to objects, qualities, and activities? How can these relations help us improve automated localization of places implicit in textual descriptions? In this paper, we motivate research on extraction of semantic place localization statements from text corpora which can be used for improving document retrieval and for reconstructing locations. The idea is to combine Semantic Web reasoning with existing geographic information retrieval (GIR) and structural text extraction for this purpose. GIR and Semantic Web technology have matured during the last years, but still largely exist in parallel. Current localization approaches have been focusing on the extraction of unstructured word lists from texts, including toponyms and geographic features, not on human place descriptions on a sentence level.

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing*Language parsing and understanding*; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods*Frames and scripts*

General Terms

Languages, algorithms

Keywords

Semantic place descriptions, text corpora, place localization, text extraction

*This is a *statement of research interest.*

ACM SIGSPATIAL COMP'13, November 5, 2013. Orlando, FL, USA

1. MOTIVATION AND RESEARCH IDEA

How can we describe places based on their *spatial and semantic relationships to objects, qualities, and activities*? How can these relations help us improve localization of and search for places implicit in textual descriptions?

Humans use many different strategies to refer to places and to localize them, for example by describing what can be done at a place, how a place was encountered during a journey, or where a place is located relative to other, known, referents. Such strategies are documented in texts about places, as well as in human speech interaction [17]. Tuan [18] has argued that an important human strategy for establishing places are *narratives*, as the latter localize places temporally as well as spatially, relative to a narrator. If it was possible to *reproduce these localization strategies in a computer*, then this would enrich place queries and improve automated place localization.

For example, historic places, such as medieval market places or land parcels, are described in historic legal documents, such as title deeds [14]. Even if these places do not exist anymore and exact locations are unknown, it is still possible to reconstruct their approximate location, based on descriptions of where they were located relative to other known places, or of what could be done there. For example, the location of land parcels may be described by relative spatial relations to surviving buildings, such as churches. And market places may be localized based on assessing the possibility to offer goods [16]. Another example are *places in a geographic landscape* which are described in mountaineering texts:

> "Piz Grisch im Winter[1]
> Einmal auf dem Grat, nimmt die Besteigung einen anderen Charakter an. [..] Besonders nach der Westseite blickt das Auge ungehemmt direkt auf den Sut-Foina Gletscher. [...] In geradliniger Stufenleiter erreichen wir den Steinmann. [...] Schlendernden Schrittes bummeln wir über nasse Wiesen zum Dörfchen hinab, [...]. Gestern haben wir auf der Höhe des Piz Grisch den Winter gefeiert, in Cresta haben wir nun den Ort gefunden, wo sich für dieses Jahr am schönsten Abschied nehmen läßt. [19]"

This German text contains references to known named pla-

[1] "Piz Grisch in winter. Once on the ridge, the ascent takes on a different character. [..] Especially towards the west side, the eye uninhibitedly looks at the Sut-Foina glacier. [...] Climbing up a linear stepladder, we reach the summit cairn. [...] We stroll through wet meadows down to the village, [...]. Yesterday, we celebrated winter at the height of the Piz Grisch, in Cresta we have now found the best place to say farewell for this year."

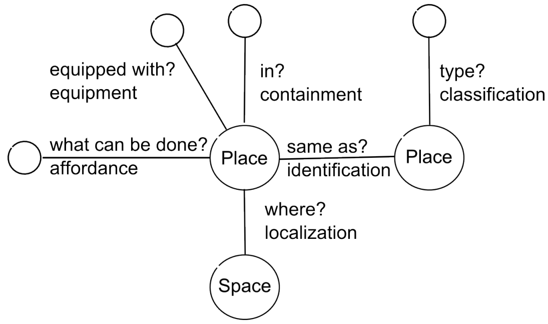

Figure 1: Essential kinds of place inference [16].

ces (Piz Grisch, Sut-Foina Gletscher, Cresta), landscape features (Grat, Wiese, Steinmann, Dorf), as well as their relation to actions (besteigen, blicken, erreichen, bummeln, finden), which can be used to describe and reconstruct the route taken by the mountaineers and to localize the places visited by them. For example, the fact that Sut-Foina is visible from somewhere on a ridge is a *localization statement*. Localization statements can be formally encoded in RDF[2] triples, e.g.:

Sut-Foina visibleFrom ?x.

, and can then be published as *linked data* [2]. In this statement, $?x$ is a variable standing for some unknown place, and *Sut-Foina* is a logical constant linked to $?x$ via the property *visibleFrom*. This description is semantically rich, which allows to relate the document to other places described in the Web and to infer the approximate location of $?x$.

The underlying relations, such as the one between places visited by the mountaineer and visible glaciers, or the one between churches and parcels, can be regarded as logical statements which can be exploited using semantic technology. It then becomes possible to search for places in documents, based on their relative locations or afforded actions. This would improve accessibility of texts in spatial and temporal search.

Some major place inference tasks relevant for our purpose are listed in the following (compare [16] and Figure 1):

- *Place equipment.* Places are associated with a certain layout and are equipped with certain landmark objects. This does not mean that these objects are part of the place or need actually be "in'" that place.

- *Place affordance.* Places afford humans to do something. This makes them primary objects of their interest. Furthermore, they can be localized indirectly via their affordances.

- *Place localization.* Places need to be localizable in space. Furthermore, we need to be able to decide *where a place is located at a given time.* The temporal parameter of place localization is essential since places can cease to exist or may even move in space [15].

These inference tasks build on each other. As we have argued in [16], places can e.g. be localized relative to *referents*, such as landmarks and ground surfaces as part of the place's equipment, *which are involved into activities afforded* by the place (see Figure 2). Similar to environmental media [15], places can be localized indirectly, based on first localizing possible activities (affordances) relative to involved referents. We will give an illustration of this approach in Section 4.

[2]Resource description framework, see `http://www.w3.org/RDF/`.

2. RELATED WORK

So far, places are mostly represented as collections of typed place names (toponyms) with footprints, i.e., gazetteers [8]. A semantic gazetteer [10], which could explicitly specify formal relations of places, remains future work. Localization strategies have often focused on unstructured *bag-of-words methods* based on tags [1, 13], or on *machine learning* and analyzing movement tracks [7]. This leaves open the question of how humans localize places in the first place, and how place equipment and affordances are related to a place. Place descriptions in texts are a rich source which can be used for this purpose. However, the challenge lies in making the semantic relations formally explicit so that machines can use it for search and localization. Semantic technology has not, to date, been used for this specific purpose.

There is work on *place descriptions* and involved concepts, on which one could built. For example, Edwardes and Purves analyzed and collected place concepts used in image tags [4], Derungs did the same for mountaineering texts [3], and Overell et al. [12] proposed a method to identify place descriptions in texts. Furthermore, several *place ontologies* have been proposed so far in the context of linked data[3].

These approaches either leave place localization to the user of a system, or do not focus on *human place localization strategies*. Furthermore, they do not make formally explicit the relations between places and other referents. Existing research on spatial reasoning has been focusing on navigation tasks, not on place identification [21].

What seems missing so far is technology which allows us to explicitly represent complex place descriptions generated by humans, to query for them instead of simple word lists, and to automatically infer localizations based on such descriptions. Places lack precise reference systems which go beyond gazetteers [16]. This hampers their use and represents a genuine research challenge in the use of Web resources for GIR. Toponym mining has typically focused on nouns, not on verbs, and thus disregards some of the relational semantics contained in a text. Semantic place descriptions directly address this research gap.

[3]See e.g. `http://vocab.org/places/schema.html`, `http://www.geonames.org/ontology`

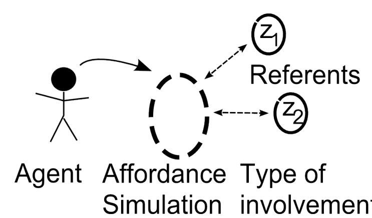

Figure 2: Involvement of agents and referents into possible activities [16].

3. EXTRACTION OF AND REASONING WITH SEMANTIC PLACE DESCRIPTIONS

We require technology that allows places to be described and localized in an easy and semi-automated manner. We suggest to address this goal by studying how places are actually described and interpreted in large thematic text corpora rich in descriptions of places, covering a theme such as mountaineering, and by investigating and reusing extraction methods which are able to cope with structural descriptions on the sentence level. We are particularly interested in *human strategies to localize and describe places*, and in technology which enables to extract this knowledge from texts and to make it formally explicit for place localization and retrieval. For this purpose, the following questions need to be answered:

1. Which geographic models of place and which concepts (classes, terms, relations) are needed for relational place descriptions? Firstly, a number of critical *queries and inferences* which need to be answerable by such concepts (*competency questions*) should be identified, following the engineering approach in [5]. Secondly, since the use of place concepts in human place descriptions is primarily an *empirical question*, the selection of concepts needs to be based on *empirical methods*. For this purpose, one could study a selection of place descriptions from some large text corpus[4]. One could reuse existing inductive geographic place models, such as [3, 13]. These models would allow to extract relevant feature classes as well as toponyms from large text corpora, including objects, activities and place qualities.

2. How can relational place descriptions (at the sentence level) be extracted from text corpora? Besides useful GIR techniques for extracting toponyms and features, such as [4], one also requires methods for extracting *part-of-speech* tags and *structural syntax*, such as Subject-Predicate-Object triples. Corresponding tools are currently developed in Computational Linguistics[5].

3. Which formalisms can be used to encode place descriptions and to automate place localization and querying? The existing tractable subsets of first-order logic (FOL) which are investigated in the Semantic Web (such as description logics (DL) and rule languages, see e.g.[6]), should be evaluated against concepts and competency questions (queries) that are necessary for our purpose. Spatial reasoning engines such as SparQ [21] could be also evaluated in this respect. Furthermore, since place localizations are approximations, it is necessary to deal to some extent with *spatial uncertainty* [11].

4. How should an inference and query infrastructure for semantic place descriptions look? We propose to build this infrastructure on *linked data principles* and *Semantic Web* technology. In this approach, all knowledge is represented as a graph with URIs for nodes as well as links, which makes it readable and available for machines in the Web. Tractable spatial query languages, such as GeoSPARQL[7], need to be integrated. A possible triple store could be, for example, Parliament[8].

Evaluation of all place related information is challenging, since it requires explicit and detailed local knowledge. As a first step, we propose to develop simple baselines, and to explore whether our approach provides meaningful improvements over these. The evaluation of localization could be based on re-localizing known places and measuring the difference to their known locations. Furthermore, based on performance comparison to existing localization methods. Evaluation of retrieval could be based on precision and recall for a known set of queries and annotated documents and compared with a set of baseline methods.

4. RECONSTRUCTING THE LOCATION OF A MOUNTAINEERING PLACE

How could we logically reconstruct, e.g., the route taken by the mountaineers in the narrative of Section 1, together with the location of the implicit places visited by them? The idea is to isolate and encode localization statements contained the text and to represent them explicitly as follows:

Sut-Foina visibleFrom ?x.
Sut-Foina eastOf ?x.
?x on ?y.
?y is a Grat.
?z reachableFrom ?x.
?z on Piz Grisch.
Cresta reachableFrom ?z.

We can now start reconstructing the route taken by the mountaineers (Figure 3). For example, where exactly is the unknown place ?y located? We know that ?y must be on a Grat (ridge), and that it must allow to see the Sut-Foina glacier towards the west. Furthermore, it must be possible to reach Piz Grisch from there. The possible locations of ?y can thus be restricted to points on a Grat from which Sut-Foina is visible towards the west. This immensely reduces the set of possible locations, if we describe them based on a terrain surface and a visibility model with respect to the Sut-Foina. If we furthermore take into account reachability (e.g., based on distances which allow to be walked) to Piz Grisch, we can restrict the set of possible locations to a small part of a single Grat which leads to Piz Grisch (compare the start of the red arrow in Figure 3).

5. DISCUSSION AND CHALLENGES

Place narratives are semantically very rich and serve as a means for establishing places [18]. We propose to exploit this richness by extracting place localization statements and making them explicit in terms of semantic descriptions. This would improve place-based search, because it allows to take into account *relations* instead of word lists, as well as place localization, because one can make use of *relative localizations* with respect to known referents.

[4]For example, landscape literature and mountaineering periodicals such as "Die Alpen. Monatsschrift des Schweizer Alpenclubs" [19], or legal text sources which describe administrative changes or land parcels [14].
[5]Zurich Dependency Parser (Parzu), available free here `https://github.com/rsennrich/parzu`.
[6]`http://www.w3.org/Submission/owl11-tractable/`
[7]`http://www.opengeospatial.org/standards/geosparql`
[8]`http://parliament.semwebcentral.org/`

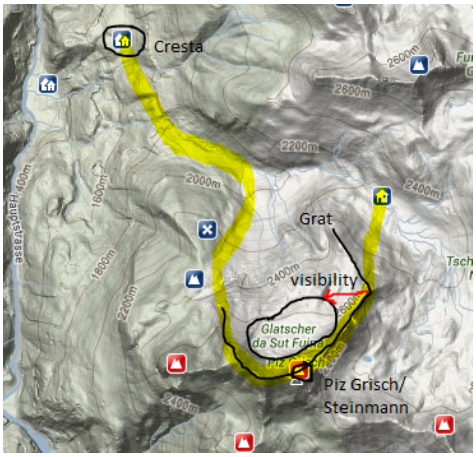

Figure 3: Places visited by the mountaineers (in yellow), reconstructed with respect to landscape features (in black) as semantic referents, and localized based on activity relations (in red).

However, explicitly describing the semantics of place narratives is a big challenge and can only be done approximately. First, as Winter and Freksa argue [20], the location of a named place, such as the Federation Square in Melbourne, may change with the context of a discourse and depending on the granularity of places with which it is contrasted. This calls for *context dependent semantic description* [9]. However, it is largely unclear how this could be solved technically. It would mean that a place may have different semantic descriptions under different discourse contexts. But how could we find out which discourse context is relevant if we search for a place? Second, the automated extraction of *explicit semantic statements*, instead of a bag of words, is an ongoing research challenge. It is by no means clear whether current methods are able to identify and filter such logical localization statements from texts in a quality which is suitable for reasoning, taking into account the complexity and ambiguity of natural speech. And third, there is a challenge related to the computation of automated localization. Higher order logic (HOL) [16] and first-order logic (FOL) are not tractable and need to be either translated to a tractable subset or amended by algorithms for computing a set of purpose-specific inferences. Semantic and spatial reasoning for place localization need to be integrated.

6. REFERENCES

[1] A. Alazzawi, A. Abdelmoty, and C. Jones. What can i do there? Towards the automatic discovery of place-related services and activities. *International Journal of Geographical Information Science*, 26(2):345–364, 2012.

[2] T. B.-L. C. Bizer, T. Heath. Linked data. the story so far. *Int. J. Semantic Web Inf. Syst.*, 5(3):1–22, 2009.

[3] C. Derungs and R. S. Purves. From text to landscape: Locating, identifying and mapping the use of landscape features in a Swiss Alpine corpus. *International Journal of Geographical Information Science*, 0(0):1–22, 2013.

[4] A. J. Edwardes and R. S. Purves. A theoretical grounding for semantic descriptions of place. In *Proceedings of the 7th international conference on Web and wireless geographical information systems*, W2GIS'07, pages 106–120, Berlin, Heidelberg, 2007. Springer-Verlag.

[5] A. Gangemi and V. Presutti. Towards a pattern science for the semantic web. *Semantic Web*, 1(1-2):61–68, 2010.

[6] J. Hightower, S. Consolvo, A. LaMarca, I. Smith, and J. Hughes. Learning and recognizing the places we go. In M. Beigl, S. Intille, J. Rekimoto, and H. Tokuda, editors, *UbiComp 2005: Ubiquitous Computing*, volume 3660 of *Lecture Notes in Computer Science*, pages 903–903. Springer Berlin / Heidelberg, 2005.

[7] L. Hill. *Georeferencing. The geographic associations of information.* MIT Press, Cambridge, MA, 2006.

[8] K. Janowicz. The role of space and time for knowledge organization on the semantic web. *Semantic Web Journal*, 1(1-2):25–32, 2010.

[9] C. Keßler, K. Janowicz, and M. Bishr. An agenda for the next generation gazetteer: Geographic information contribution and retrieval. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '09, pages 91–100, New York, NY, USA, 2009. ACM.

[10] D. Montello, M. Goodchild, J. Gottsegen, and P. Fohl. Where's downtown? Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition & Computation*, 2(3):185–204, 2003.

[11] S. Overell and S. M. Rüger. Identifying and grounding descriptions of places. In R. Purves and C. Jones, editors, *GIR*. Department of Geography, University of Zurich, 2006.

[12] R. S. Purves, A. J. Edwardes, and J. Wood. Describing place through user generated content. *First Monday*, 16(9), 2011.

[13] S. Rees-Jones. Database of medieval title deeds for the city of york. *Occasional Papers in History*, (3), 1996.

[14] S. Scheider and K. Janowicz. Places as media of containment. In *Extended Abstracts Volume, GIScience 2010*. University Zürich-Irchel, Zürich, 2010.

[15] S. Scheider and K. Janowicz. Place reference systems. A constructive activity model of reference to place, 2013. `http://www.geographicknowledge.de/pdf/PlaceReferenceSystems.pdf`.

[16] Y. Tuan. *Space and place: The perspective of experience.* University of Minnesota Press, Minneapolis, 1977.

[17] Y.-F. Tuan. Language and the making of place: A narrative-descriptive approach. *Annals of the Association of American Geographers*, 81(4):684–696, 1991.

[18] E. Wenzel. Piz grisch im winter. In *Die Alpen. Monatsschrift des schweizer Alpenclubs*, pages 139–142. Stämpfli, 1940.

[19] S. Winter and C. Freksa. Approaching the notion of place by contrast. *J. Spatial Information Science*, 5(1):31–50, 2012.

[20] D. Wolter and J. O. Wallgrün. *Qualitative Spatial Reasoning for Applications: New Challenges and the SparQ Toolbox.* IGI Global, 2010.