Deriving place graphs from spatial databases

Ehsan Hamzei
Dept. Infrastructure Engineering
University of Melbourne
ehamzei@student.unimelb.edu.au

Hao Chen
Dept. Infrastructure Engineering
University of Melbourne
hchen@student.unimelb.edu.au

Hua Hua
Research School of Computer Science
Australian National University
hua.hua@anu.edu.au

Maria Vasardani
Dept. Infrastructure Engineering
University of Melbourne
maria.vasardani@unimelb.edu.au

Martin Tomko
Dept. Infrastructure Engineering
University of Melbourne
tomkom@unimelb.edu.au

Stephan Winter
Dept. Infrastructure Engineering
University of Melbourne
winter@unimelb.edu.au

Abstract

A place graph is an abstract representation of human place knowledge, which models spatial references. A place graph can be used for various tasks that rely on reasoning and querying of the stored knowledge. In related work, place graphs were constructed from parsing natural language place descriptions using language processing techniques. In this research, we present an innovative approach to derive place graphs from information stored in spatial databases, with a demonstration using OpenStreetMap data. The approach provides a complementary way to generating place graphs from natural language descriptions.

1 Introduction

Place graphs are spatial property graphs designed to model *spatial references.* Spatial references locate places (nodes) by their spatial relationships (directed edges) to other places, e.g., "The *courtyard* is on the *campus*, beside the *clocktower*" describes the location of the courtyard in relation to the campus and the clocktower. Spatial references can be extracted from natural language (NL) place descriptions, for example, (Vasardani et al., 2013), as these are used to communicate spatial information about places. The extracted spatial references are in the form of *triplets*, each of a *locatum* (L), a *relatum* (R), and the qualitative spatial relationship between them (r): for the example <L: courtyard, r: on, R: campus> and <L: courtyard, r: beside, R: clocktower>. Triplets can be extracted from place descriptions using NL processing technologies (Vasardani et al., 2013; Liu et al., 2014). Figure 1 illustrates the place graph generated from the two examples above.

Such a place graph forms an abstract representation of spatial knowledge about place and has been used for tasks such as reasoning (Chen et al., 2015), georeferencing (Chen et al., 2017), sketch map drawing (Kim et al., 2016a), and extracting local landmarks (Kim et al., 2016b). It has also been hypothesized to be a suitable knowledge base supporting place-based querying and question answering. Since common language place

Figure 1: The place graph representing the spatial references “the courtyard is on the campus” and “the courtyard is beside the clocktower”.

descriptions can be difficult to collect for all desired environments, we seek alternative ways to generate place graphs.

As shown in Figure 2, a GIS, a place description, and a place graph are alternative representations of place knowledge. Most current approaches derive place graphs from NL descriptions and link places in the graphs to a GIS through georeferencing, as indicated by the three solid arrows. This research focuses on one of the missing links, as shown by the dashed arrows i.e., deriving place graphs from GIS. In the future, this can be combined with techniques to generate NL place descriptions from place graphs, thus resolving generating place descriptions from GIS. Completing the circle can smooth and simplify human-computer interaction in terms of translating place knowledge among different representations.

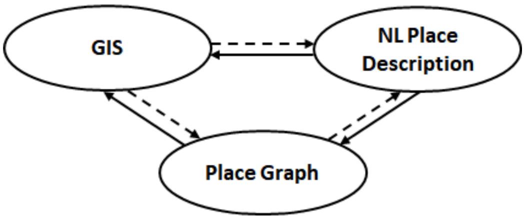


Figure 2: Translation among three different representations of spatial knowledge about place, with achieved links indicated by solid arrows.

This research will investigate methods to derive place graphs from spatial databases (e.g., a GIS), where places are represented by features with either point, polyline, or polygon geometry. Accordingly, the hypothesis of this paper is that place graphs can be generated from information stored in spatial databases. The task is not trivial although computing qualitative spatial relationships (in the approach below, topological and directional ones) is straight-forward; the real challenge is to derive place graphs in a way that is similar to how people cognitively represent and communicate about place and their spatial relationships.

The rest of this paper is structured as follows: In Section 2 we review related work. In Section 3 our approach for deriving place graphs from spatial databases is presented. Section 4 explains the implementation and a case study using OpenStreetMap data, as well as a discussion of the obtained results. Section 5 concludes this work.

2 Related Work

Place based research is an emerging field in GIScience, and its importance has been widely acknowledged (Goodchild, 2007, 2011; Winter et al., 2016). The goal is to facilitate human-computer interactions through modeling and utilizing place-related information. For example, Egenhofer and Mark (1995) suggested the term *Naïve Geography* to capture and reflect the way that non-expert think and reason about space and time.

Vasardani *et al.* (2013) focus on locative expressions — parts of NL descriptions that provide location information of a place reference using a spatial relation to another place as landmark, e.g., *“the lawn is in front of the library”*. Each locative expression can be modeled by a triplet, and triplets can be extracted automatically from place descriptions using existing NL parsers (Liu et al., 2014). Place graphs can then be constructed from such triplets automatically (Kim et al., 2016a). Each triplet is stored by two nodes in a place graph, one each for the locatum and the relatum, and an edge for the spatial relation. A place graph can be constructed from multiple descriptions, in which case place references identified for the same place are stored in one node (merged) (Kim et al., 2016c). Thus, the place graph can be considered as a model of collective human place knowledge extracted from NL descriptions (Kim et al., 2016b). Compared to the object- (e.g., in a gazetteer) and field-based models (e.g., (Jones et al., 2008)) of places, the place graph additionally captures the *network* dimension (Kuhn, 2012) of place information. Place graphs have been used for various tasks such as georeferencing (Chen et al., 2017), or landmark extraction (Kim et al., 2016b) in previous studies.

This research is informed by human reasoning and communication about places and spatial relationships. Richter *et al.* (2013) collected and analyzed NL place descriptions and found that people tend to describe places in a hierarchical manner — a property that has also been observed in the organization of cognitive maps (Hirtle and Jonides, 1985). We capture such hierarchical structures using containment relationships between places. Our approach also considers directional relationships, another type of qualitative spatial relationships widely studied in spatial cognition and AI (e.g., Freksa (1992); Frank (1992)). Miller (1956) convincingly argued that human short-term memory capacity is limited and approximately bounded to seven plus/minus two units of information. Hence, in this study, this limitation is used as a threshold for grouping places before investigating their qualitative spatial relationships.

3 Approach

Figure 3 shows the sequence of processing for creating place graphs from data stored in spatial databases. The proposed approach consists of three steps: (1) the extraction of a hierarchical structure based on containment relationships, (2) updating the hierarchy using a quad-tree strategy, and (3) the extraction of qualitative spatial relationships from quantitative data. Up till now, place graphs were the result of processed NL descriptions. The method suggested here is built upon cognitive studies in order to produce place graphs that resemble those from people's descriptions, i.e., not necessarily complete, but with relatively few, salient relationships between places. Compared to a complete, fully connected graph, the produced place graph is smaller in storage size (thus allowing for faster querying) and better captures the way people tend to describe places in a given environment. It is worth noting that using qualitative spatial reasoning non-stored relationships can be inferred.

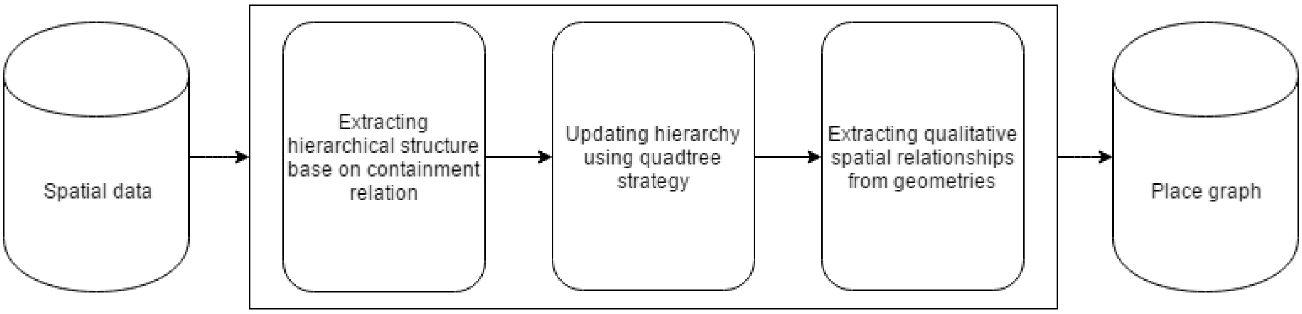


Figure 3: Proposed approach for creating place graphs using spatial data.

3.1 Extracting hierarchical structure based on containment relationships

The first step generates a hierarchical structure, represented by containment relationships between places. The hierarchy is created in two sequential steps using Algorithm 1. First, the containment relationships between each pair of polygons are checked, to form a containment network. Then, the containment network is pruned into a hierarchical structure.

3.2 Updating the hierarchy using a quad-tree strategy

After creating the hierarchical structure, a quad-tree strategy is applied to update the hierarchy by creating intermediate nodes. The idea behind these intermediate nodes is cognitively motivated, based on the limited human short-term memory span. As mentioned earlier, Miller (1956) argues that human capacity for processing information is limited to approximately seven plus/minus two elements. Hence, if a node contains more than a fixed number of nodes (five in this research), the polygon associated with the node is divided into four polygons based on its centroid, using a quad-tree. These polygons are linked to intermediate nodes that are placed between a node and its contained nodes in the updated hierarchical structure. Then, the containment relationships between the contained nodes and intermediate nodes are checked. Finally, the relationships will be pruned to maintain the hierarchical structure. This process iteratively updates the hierarchy to a new version until each node, including intermediate nodes, has no more than five contained child nodes in the hierarchy. Algorithm 2 shows how the quad-tree strategy is used in updating the hierarchy.

Algorithm 1 generating the hierarchical structure

```
procedure Hierarchy(polygons)
    create a hierarchy node for each polygon (p)
    for polygon (p) in the polygons do
        for other polygon (o) in the polygons do
            if p contains o then
                create a containment relationship between their nodes (p contains o)
            end if
        end for
    end for
    for every node (n) do
        for every node (c) which is contained in n do
            if c contains any other node o then
                remove the relationship between n and o
            end if
        end for
    end for
end procedure
```

Algorithm 2 updating hierarchical structure

```
procedure Update(nodes)
    for every node n from the root to the leaves do
        if number of contained nodes c is more than the predefined threshold then
            create four polygons by dividing polygon n (north-west, north-east, south-west, south-east)
            create four nodes (r) in the hierarchy for each of the created polygons
            create containment relationship between n and new nodes r
            remove relationships between c and n
            create containment relationships between nodes r and c
        end if
    end for
end procedure
```

3.3 Extracting qualitative spatial relationships from quantitative data

In this step, the updated hierarchical structure is used for further populating topological and directional relationships. For each pair of sibling nodes in the hierarchy, their cardinal direction relationship, e.g., ‘north’ or ‘east’, and topological relationship, e.g., ‘overlap’ or ‘meet’ (Egenhofer and Franzosa, 1991), are determined, based on their geometries. Finally, a place graph is created with nodes from the hierarchy and the derived relationships. Algorithm 3 shows the process from the hierarchical structure to the place graph.

4 Experimental Results

4.1 Implementation

The proposed approach is implemented as a toolbox for creating place graphs using OpenStreetMap (OSM) data. Figure 4 shows the workflow of creating place graphs from OSM maps. First, by defining the target area and using OSM API, the input data is gathered in XML format. Then, by parsing the XML file, polygons are extracted for further analysis. Next, the previously described approach is used to generate a place graph from the extracted polygons (in this research we do not consider places with polyline or point geometries). Finally, the results are stored in a graph-based place database. In this study, Neo4j Community version and Gephi are used as the graph-based database and graph visualizer, respectively.

4.2 Experiments

Three experiments are designed to test the approach for creating place graphs in different map scales. Neighborhoods of Melbourne Cricket Ground, neighborhoods of Melbourne CBD, and the Greater Melbourne area are considered as geographic areas of interest. These experiments are conducted to show the effectiveness of

Algorithm 3 generating place graph

```
procedure PlaceGraph($nodes$)
    create empty graph structure
    put hierarchy into the graph
    for every node ($n$) from the root to the leaves do
        for every pair of nodes ($c1$, $c2$) contained in $n$ do
            generate the appropriate cardinal direction relationship between $c1$ and $c2$
            generate the appropriate topological relationship between $c1$ and $c2$
        end for
    end for
    insert graph into graph database
end procedure
```

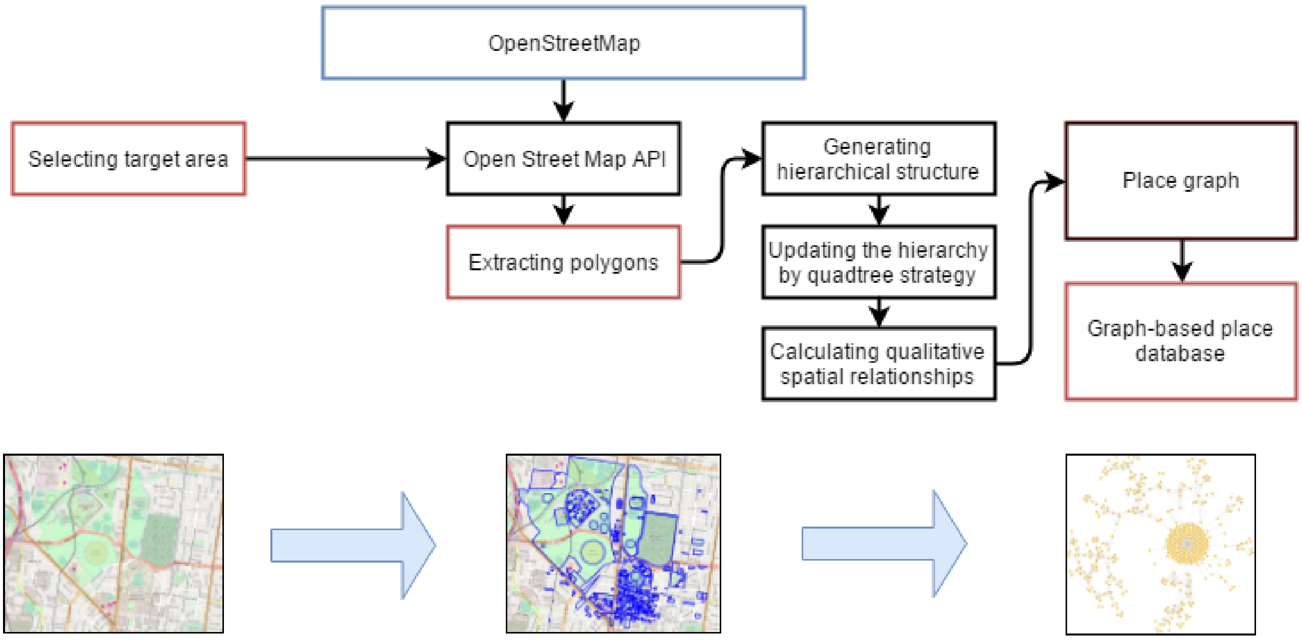
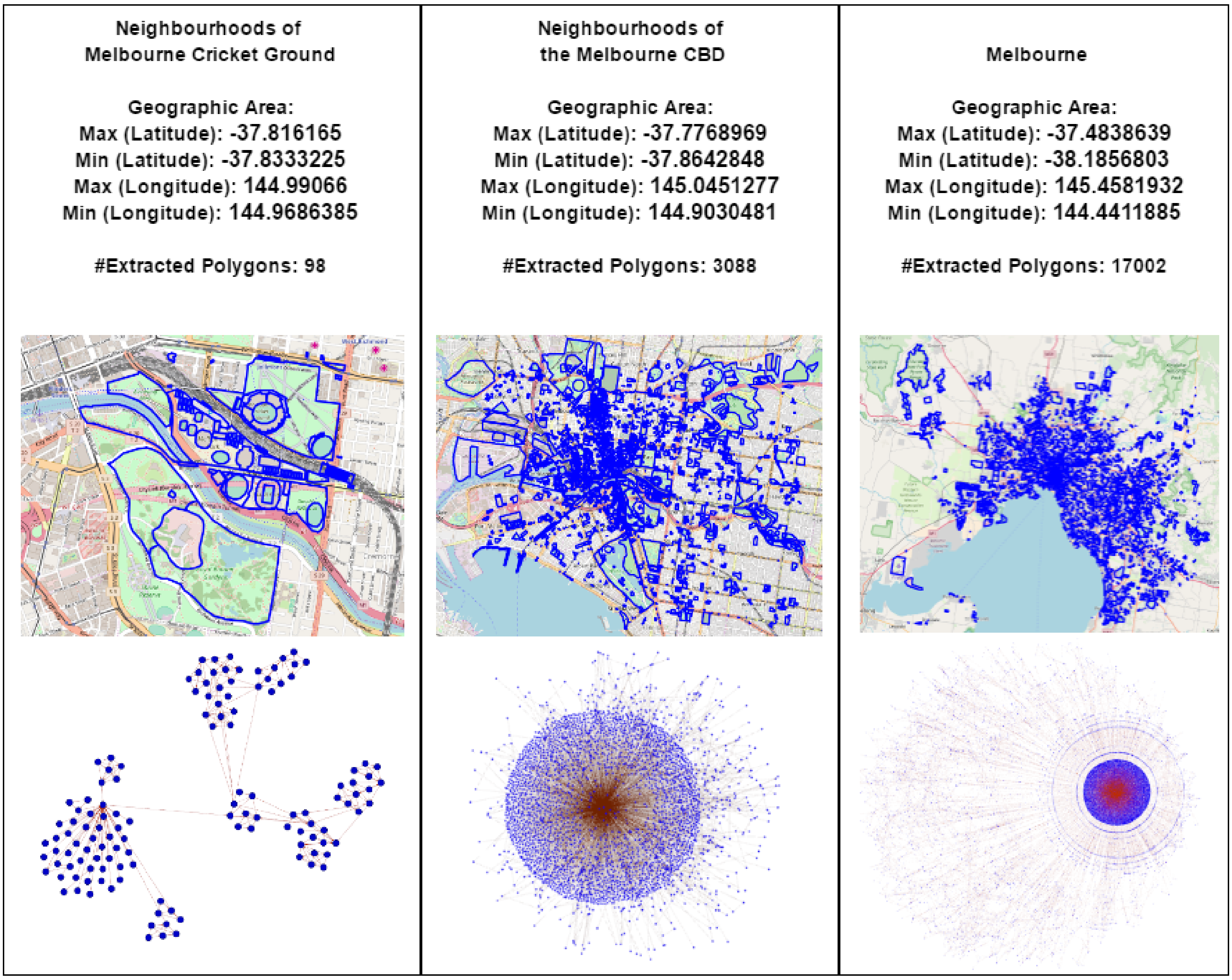


Figure 4: Workflow of generating place graphs from maps.

the proposed approach for generating place graphs at different scales. The extracted polygons are processed to produce the place graphs. The geographic areas and the produced place graphs are shown in Figure 5.

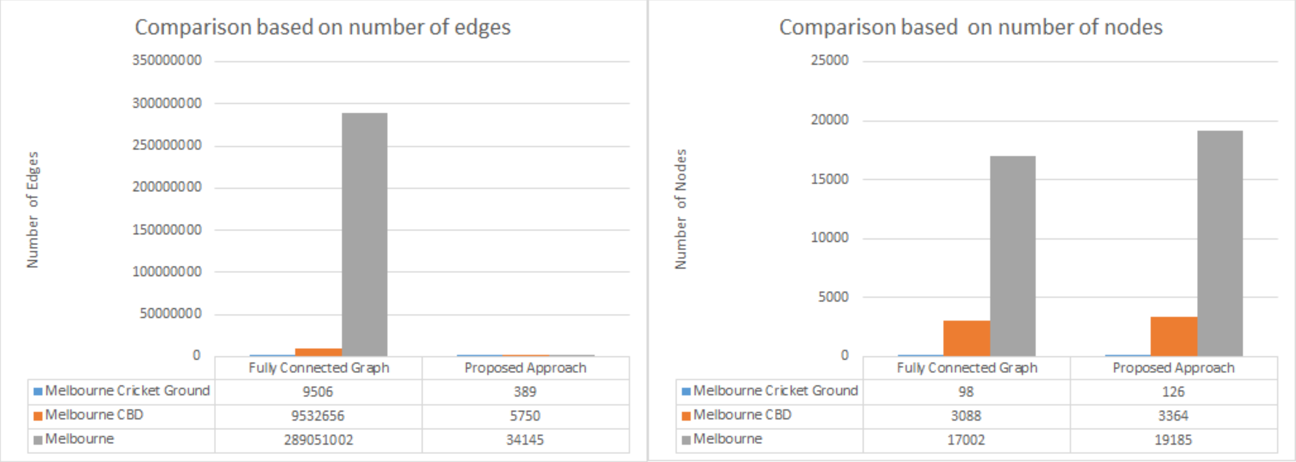
The approach is compared to a baseline method that creates fully connected graphs. In a fully connected graph, every pair of nodes is connected with two different edges, one for cardinal direction and the other for topological relationships. Equation 1 allows to determine the number of edges in a fully connected graph. The number of edges depends on how the graph is stored. Hence, equation 1 is held when topological and directional relationships are stored separately. Figure 6 shows the number of edges and nodes for the proposed approach and the fully connected graph method. The number of nodes in our method is slightly higher than the fully connected graph due to the intermediate nodes resulting from the quad-tree strategy. However, the number of edges is much lower in the proposed approach compared to the fully connected graph method. Hence, the graphs resulting from our approach require comparatively less storage. It is worth noting that there is not any information loss in the extracted place graph compared to the fully connected graph. This is because every relationship in the fully connected graph either also exists in the extracted place graph, or can be inferred using qualitative spatial reasoning. Moreover, the proposed method is based on cognitive constraints — the limitation of short-term memory (Miller, 1956) and the hierarchical structure of places in human cognitive maps (Hirtle and Jonides, 1985).

$$numberOfEdges = numberOfNodes \times (numberOfNodes - 1) \tag{1}$$



Neighbourhoods of Melbourne Cricket Ground	Neighbourhoods of the Melbourne CBD	Melbourne
Geographic Area:	Geographic Area:	Geographic Area:
Max (Latitude): -37.816165	Max (Latitude): -37.7768969	Max (Latitude): -37.4838639
Min (Latitude): -37.8333225	Min (Latitude): -37.8642848	Min (Latitude): -38.1856803
Max (Longitude): 144.99066	Max (Longitude): 145.0451277	Max (Longitude): 145.4581932
Min (Longitude): 144.9686385	Min (Longitude): 144.9030481	Min (Longitude): 144.4411885
#Extracted Polygons: 98	#Extracted Polygons: 3088	#Extracted Polygons: 17002

Figure 5: Geographic areas and results of the experiments.



	Fully Connected Graph	Proposed Approach
■ Melbourne Cricket Ground	9506	389
■ Melbourne CBD	9532656	5750
■ Melbourne	289051002	34145

	Fully Connected Graph	Proposed Approach
■ Melbourne Cricket Ground	98	126
■ Melbourne CBD	3088	3364
■ Melbourne	17002	19185

Figure 6: Comparison of proposed approach and fully connected method.

5 Conclusion

This paper proposes an innovative approach to generate place graphs from spatial databases. In the place graphs generated, containment relations are represented in a hierarchical structure, with topological and cardinal direction relations captured between sibling nodes. The approach considers various cognitive factors of how people think about and describe places, as well as their spatial relationships. It also considers the limitation of human short-term memory and the hierarchical structure of places in human cognitive maps. It is the first attempt to derive place graphs from spatial databases. In future work, we plan to consider deriving qualitative distance relationships, e.g., 'near', from spatial databases, for example based on contrast set theory (Winter and Freksa, 2012); as well as to derive relative direction relationships, e.g., 'in front of' or 'left of', by considering the different frames of reference and points of view used in place descriptions. In addition, testing the closeness of the extracted place graphs from different sources, NL place descriptions and spatial databases, can be another future work of this study.

Acknowledgements

The support by the Australian Research Council grant DP170100109 is acknowledged.

References

Hao Chen, Maria. Vasardani, and Stephan Winter. Maintaining relational consistency in a graph-based place database. In *Proceedings of Research@Locate 15*. Locate15, 2015.

Hao Chen, Maria Vasardani, and Stephan Winter. Geo-referencing place from everyday natural language descriptions. *arXiv preprint arXiv:1710.03346*, 2017.

Max J. Egenhofer and Robert D. Franzosa. Point-set topological spatial relations. *International Journal of Geographical Information Systems*, 5(2):161–174, 1991.

Max J Egenhofer and David M Mark. Naive geography. In *International Conference on Spatial Information Theory*, pages 1–15. Springer, 1995.

Andrew U Frank. Qualitative spatial reasoning about distances and directions in geographic space. *Journal of Visual Languages & Computing*, 3(4):343–371, 1992.

Christian Freksa. Using orientation information for qualitative spatial reasoning. In Andrew U. Frank, Irene Campari, and Ubaldo Formentini, editors, *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, volume 639 of *Lecture Notes in Computer Science*, pages 162–178. Springer, 1992.

Michael F. Goodchild. Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007.

Michael F. Goodchild. Formalizing place in geographical information systems. In L. M. Burton, S. P. Kemp, M.-C. Leung, S. A. Matthews, and D. T. Takeuchi, editors, *Communities, Neighborhoods, and Health: Expanding the Boundaries of Place*, pages 21–35, New York, 2011. Springer.

Stephen C Hirtle and John Jonides. Evidence of hierarchies in cognitive maps. *Memory & Cognition*, 13(3): 208–217, 1985.

Christopher B Jones, Ross S Purves, Paul D Clough, and Hideo Joho. Modelling vague places with knowledge from the web. *International Journal of Geographical Information Science*, 22(10):1045–1065, 2008.

Junchul Kim, Maria Vasardani, and Stephan Winter. From descriptions to depictions: A dynamic sketch map drawing strategy. *Spatial Cognition & Computation*, 16(1):29–53, 2016a.

Junchul Kim, Maria Vasardani, and Stephan Winter. Landmark extraction from web-harvested place descriptions. *KI-Künstliche Intelligenz*, 2(31):151–159, 2016b.

Junchul Kim, Maria Vasardani, and Stephan Winter. Similarity matching for integrating spatial information extracted from place descriptions. *International Journal of Geographical Information Science*, 1:1–25, 2016c.

Werner Kuhn. Core concepts of spatial information for transdisciplinary research. *International Journal of Geographical Information Science*, 26(12):2267–2276, 2012.

Fei Liu, Maria Vasardani, and Timothy Baldwin. Automatic identification of locative expressions from social media text: A comparative analysis. In Dirk Ahlers, Erik Wilde, and Bruno Martins, editors, *4th International Workshop on Location and the Web*, pages 9–16. ACM, 2014. ISBN 978-1-4503-1459-6.

George A Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2):81, 1956.

Daniela Richter, Stephan Winter, Kai-Florian Richter, and Lesley Stirling. Granularity of locations referred to by place descriptions. *Computers, Environment and Urban Systems*, 41:88–99, 2013.

Maria Vasardani, Sabine Timpf, Stephan Winter, and Martin Tomko. From descriptions to depictions: A conceptual framework. In Thora Tenbrink, John G. Stell, Antony Galton, and Zena Wood, editors, *Spatial Information Theory*, volume 8116 of *Lecture Notes in Computer Science*, pages 299–319. Springer, 2013. ISBN 978-3-319-01789-1.

Stephan Winter and Christian Freksa. Approaching the notion of place by contrast. *Journal of Spatial Information Science*, 5(1):31–50, 2012.

Stephan Winter, Timothy Baldwin, Jochen Renz, Martin Tomko, and Werner Kuhn. Place knowledge as a trans-disciplinary research challenge for geographic information science. In Jeremy Mennis, editor, *UCGIS Symposium*, 2016.