

Article

A Graph Database Model for Knowledge Extracted from Place Descriptions

Hao Chen ^{1,*} , Maria Vasardani ¹ , Stephan Winter ¹  and Martin Tomko ¹ 

¹ University of Melbourne, Parkville, Victoria, 3010, Australia

* Author to whom correspondence should be addressed.

* Correspondence email: hchen@student.unimelb.edu.au

Abstract: Everyday place descriptions provide a rich source of knowledge about places and their relative locations. This research proposes a place graph model for modeling this spatial, non-spatial, and contextual knowledge from place descriptions. The model extends a prior place graph, and overcomes a number of limitations. The model is implemented using the Neo4j graph database, and a management system has also been developed that allows operations including querying, mapping, and visualizing the stored knowledge in an extended place graph. Then three experimental tasks, namely georeferencing, reasoning, and querying, are selected to demonstrate the superiority of the extended model.

Keywords: place; spatial property graph; place graph; graph database; place description

1. Introduction

Place descriptions occur in everyday verbal communication as a way of encoding and transmitting spatial and semantic knowledge about places between individuals [1]. They also occur in written forms such as news articles, social media texts, trip guides, and tourism articles, and thus the web provides a plethora of place descriptions as well. It has been postulated that knowledge conveyed by place descriptions might form a place-based information system [2], in contrast to space-based GIS. In order to set up such systems, three relevant research problems can be identified: information extraction through natural language (NL) processing, information modelling, and knowledge creation and utilisation. For the first task, techniques such as gazetteered place name identification (e.g., [3–5]) and spatial relationship extraction (e.g., [6–8]) from texts have already been developed. For the second task a prior graph-based model exists [1] that will be significantly extended in this research. Accordingly, this research focuses on the second and third tasks, i.e., modeling and utilizing knowledge extracted from place descriptions.

Place descriptions typically provide a qualitative reference system for describing geographic locations, and consist essentially of references to places and their qualitative spatial relationships. Information extracted from place descriptions is used to construct *place graphs* [1], which consist of places as nodes and spatial relationships as edges. The edges are directed from locatum (L) to relatum (R). For example, the description ‘The *courtyard* is on the *campus*, beside the *clocktower*’ describes the location of the courtyard in relation to two other places, the campus (as a container) and the clocktower (as a neighbour). This information can be modelled in the form of *triplets* of a *locatum* *L* – the reference to a place that is to be located, a *relatum* *R* – the reference to a place that is already located, and a spatial relationship *r* between the two: $\langle L: \text{courtyard}, r: \text{on}, R: \text{campus} \rangle$ and $\langle L: \text{courtyard}, r: \text{beside}, R: \text{clocktower} \rangle$. The two triplets are used to construct the simple place graph shown in Figure 1.



Figure 1. Place graph consisting of two example triplets: $\langle \text{courtyard}, \text{on}, \text{campus} \rangle$ and $\langle \text{courtyard}, \text{beside}, \text{clocktower} \rangle$.

Though these place graphs are place knowledge bases [9,10], the triplets that build them are stripped off of much of their conversational contexts. It is, therefore, possible to find incompatible information in a place graph, especially if the graph is constructed from combining place descriptions with different contexts. For example, it is perfectly possible to collect seemingly contradicting triplets such as <Melbourne, near, Sydney> and <Melbourne, far, Sydney> from two place descriptions, since the use of the different distance prepositions is context-dependent. Storing such triplets in a place graph without further capturing their original contexts could result in loss of information and misinterpretation. Moreover, the interpretation of spatial relations found in NL place descriptions often relies on information that is not explicit and has to be inferred. For example, people infer whether the reference frame of relative direction relations, as in <courtyard, left of, clocktower> is intrinsic or relative. While people are often capable of such inferences, the basic place graph only stores explicit information. The above place graph also do not consider place semantics or place-related human activities, which are valuable for further spatial reasoning and query answering.

Consequently, the usefulness of the original place graph model from a knowledge base perspective is restricted. Currently, query answers are provided by matching the values of certain property keys and by graph traversing, without filtering for context, inferences, or semantics, while the interpretation of spatial relationships such as relative directions or qualitative distances is limited.

This work reorganizes, revises and extends the original place graph model. The goal is to capture information from place descriptions that is useful, but lost during the triplet extraction. The hypothesis of this research is that, the extended place graph overcomes significant limitations of the original model in georeferencing, reasoning, and querying tasks. Specifically, the extended model can be used to derive more constrained locations of places for georeferencing. It captures additional information such as reference frames to be used in maintaining relational consistency, and it is capable of answering additional spatial queries. Accordingly, the contributions of this research include:

1. The identification of eight types of information that are either embedded in place descriptions or in external context and have not been captured by the original place graph model.
2. An extended place graph model that represents such information and enables future tracing as well as querying.
3. The implementation of the extended place graph into a graph database management system, which allows operations including querying, visualizing, and mapping.
4. The demonstration of how the extended model overcomes limitations of the original place graph in georeferencing, reasoning, and querying tasks based on three experiments.

The remainder of this paper is structured as follows: In Section 2 related work on place, place models, place descriptions, and place graphs is provided. Section 3 identifies the information not captured in the original place graph model, and introduces the extended model. Section 4 looks at the extended model's implementation and three experiments demonstrating its superiority. In Section 5 the experimental results are discussed, and the highlights of this work are presented in the concluding Section 6.

2. Related Work

Place based research is an emerging field in GIScience with importance widely acknowledged (e.g., [11–13]). The purpose is to smooth and simplify human-computer interaction by capturing, modelling, and utilizing place-related information. For example, Egenhofer and Mark suggested *Naive Geography* in order to capture and reflect the way that non-experts think and reason about space and time [14]. In this section, related work about how places are conceptualized, modelled in information systems, and communicated in descriptions is discussed.

2.1. Place as a cognitive concept

Space and place are two fundamental concepts in geography, and more broadly in social sciences, humanities, and information science [15]. Although the concept of place has existed for

long in philosophy and psychology [16], it is relatively new in the geographic information domain. People talk about space by referring to places [17], and the definition of place has been discussed extensively (e.g., [18,19]). Compared to a space-based perspective of geography, place is regarded as space infused with human meaning and experience, and thus, enables conversations [20]. Place is also regarded by some as the prototypical spatial reference in human, economic, and culture geography [21].

Place, as a cognitive concept, is inherently vague, and this vagueness is evident in human cognition, perception, as well as natural language descriptions [22]. Some researchers believe the concept of place may be too vague to be formalized, except in narrow circumstances [12]. It has been argued that places do not have any natural boundaries, and are locations that have been given shape and form by people [19,23]. Agnew [24] suggests thinking of place in relation to other places, instead of bounded and isolated features. In contrast, geographic information systems (GIS) and services are developed on unambiguous, crisp, and metric geometries removed from human concepts [25], and has successive therefore limited ability in modeling and utilizing place information [26]. In short, place, while fundamental to human cognition and communication, is still well beyond the reach of current information systems.

2.2. Place models from an information system perspective

Web-based place services such as Google typically use gazetteers to store the locations of places by point. A gazetteer, which is a dictionary of geographic names, contains three core components: place names, feature types, and footprints [27,28]. A gazetteer stores usually the official or authoritative place names, and sometimes also stores alternative names such as vernaculars. A place type is a category from a feature-type thesaurus for classifying places according to their semantics, and is often biased towards political or commercial entities and geographic features with large extents. A footprint represents the location of a place, typically by a single coordinate pair as an estimated center of a place, which is not capturing its extension and often inappropriately precise [19]. Gazetteers are widely utilized for both enterprise and academic purposes such as geographic information retrieval (GIR) [29–31], navigation services, and web-mapping applications.

Vague places such as ‘the South of France’ have been modelled using field models in some studies to represent the degrees to which any location belongs to these places. Montello and Goodchild conducted a study to determine the footprint of *downtown Santa Barbara* by asking participants to draw the boundary of it and aggregating the results [32]. Later, data-driven methods were proposed using techniques such as density analysis and clustering, based on geotagged social media content of place names or tags [33–36]. Other studies focus on deriving continuous surfaces to represent places [37,38]. Such field-based representations computationally characterize the inherent uncertainty of the extent of vague places, and enable approximate crisp boundaries to be derived for place-based applications.

Winter and Freksa argue that the spatial extent of a place referred to by a place name could vary in different contexts [19]. Places function as spatial anchors, and are determined by their relationships to other places in the environment. This idea refers to a sense of distinctiveness [39] and wholeness. Winter and Freksa suggest to capture the cognitive and linguistic nature of a place in contrast to other places that are relevant to the discourse. Such contrast sets of places can either be explicitly mentioned in the discourse, implied, or pre-exist as shared knowledge. Using contrast sets, Vasardani *et al.* provided a conceptual model to interpret the region implied by preposition *at*; the contrast sets make the results context-sensitive [40].

Non-spatial place information such as place semantics, equipment, characteristics, and affordance can also be useful in applications such as place searching and querying, and some of them have already been studied. For example, it has been argued that place affordance is a core component for defining place and designing ontologies, and several works have attempted to formalize it [41–43]. Semantical categories used in gazetteers can be regarded as taxonomies according to place affordance, although sometimes lack flexibility and interoperability [44].

2.3. Modelling place descriptions

The structure of place descriptions has been studied in linguistics and spatial cognition research (e.g., [45,46]), and recently in the domain of GIScience as well. For example, Richter *et al.* found that place descriptions typically apply hierarchical structures from different granularity levels [47]. Place descriptions provide a qualitative reference system for describing geographic locations using references to places (often as place names) and spatial relationships. Computers have difficulties recognizing and interpreting such structures, often due to vernacular place references and flexible relationship expressions [17].

Bateman *et al.* developed a comprehensive linguistic ontology for processing spatial language [48]. Some parsers for annotating references to places (or spatial objects in some studies) and spatial relationships in natural language text have already been developed [6,8,49]. Vasardani *et al.* study place references and spatial relationships embedded in locative expressions, which can be extracted by a parser in the form of triplets [1]. A place reference can be a place name (also called a toponym, e.g., 'Paris') or a count noun (typically a place category, e.g., 'the library') [18]. It can also be in other vernacular forms (e.g., 'the meeting place'). References that are not official place names are more challenging to locate and typically require considering conversational contexts. A triplet provides location information of a place by giving relationships to other places as anchors or landmarks. The spatial relationship expressed in a locative expression is often a preposition (e.g., *on*, *in* or *at*), but can also be a verb (e.g., *surrounding*), a phrase, or even be implicit (e.g., *Sydney, NSW*).

Place graphs from triplets can have multiple edges between each pair of nodes, representing different spatial relationships between the two places, because the same places may be referred to multiple times, or in several place descriptions. For each type of relationship, at most one instance is stored between any pair of nodes as an edge, and additional instances will be regarded as duplicates and discarded.

The references to the same place may also be made in different terms. Thus, nodes referring to identical places should be identified and merged. For this purpose, a graph-merging approach considering string and semantic similarity as well as similarity of spatial relationships to other places has been developed [9]. An example of a merged place graph is shown in Figure 2, constructed from multiple descriptions of the same environment. Place graphs have been leveraged to locate non-gazetteered references to places [50], create sketch maps [51], and identify landmarks [10].

3. Extending the Place Graph Model

This section first analyzes types of information that are not captured in the original place graph model, as well as the tasks for which they matter. Then, an extended place graph model that caters for this information is introduced.

3.1. Information not captured in the original place graph model

The types of information identified below are not considered in the original model. Most of them provide contextual knowledge, which could affect the interpretation of other information communicated in place descriptions (e.g., spatial relations), and, thus, should be captured. The definition of context is task-specific; in this research we adopt the categorization proposed by Wolter and Yousaf [52] of description-, environment-, and human-dependent contexts, as shown in Figure 3. For instance, *near* can refer to different distances according to other places relevant to the discourse [19] (description-dependent context). Certain relations require information from the environment in order to be interpretable, e.g., 'two blocks down the street' (environment-dependent context). Places and spatial relations can also have different semantics for different individuals (human-dependent context).

Figure 4 shows the UML diagram of the original place graph model.



Figure 2. A merged place graph constructed from descriptions of the University of Melbourne, with node size and color corresponding to node degree, and edge size and color corresponding to number of edges between nodes (left side), and details of a zoomed-in sub-graph (right side).

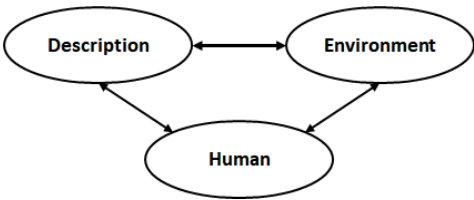


Figure 3. Contextual factors that affect the interpretation of spatial information in place descriptions.

3.1.1. Place semantics and characteristics

Place descriptions sometimes contain non-spatial information about places, such as their types (e.g., ‘the room is a lecture theater’), the activities they afford (e.g., ‘having seminars and lectures’), the things they equip (e.g., ‘the room has a projector’), as well as their characteristics (e.g., ‘old, large’) [53,54].

Place semantics and affordances have been used for characterizing places and enabling platial search as well as analysis [38,54,55]. Different places may have the same affordances, and one place may have multiple affordances according to individuals or time periods. The way that a gazetteer categorizes places does not always align with the way people regard these places, despite that such categorization is useful in many applications. Capturing semantics and characteristics of places in a place graph could provide additional dimensions for tasks such as georeferencing, identical-place matching, and querying.

In place descriptions, these types of information are often expressed in certain patterns, e.g., as adjectives, nouns followed by words such as ‘is’ and ‘has’, or as verb phrases. Such patterns can be recognized using a trained parser, and the feasibility of creating such a parser has been demonstrated in previous research (e.g., [54,56]).

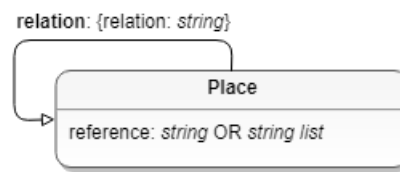


Figure 4. UML of the original place graph model.

3.1.2. Places and relationships from discourse and their sequential order of appearance

Places referred to in different discourse provide contextual knowledge for interpreting spatial relations and locating places. For instance, *near* in the description ‘the building is near the Flinders Street Station’ can be interpreted differently in terms of distance, depending on the spatial context (the geographic extent the description is embedded in), e.g., the limited area around the station, or the whole Melbourne CBD. Such a spatial context can be inferred by looking at the places mentioned in the same discourse.

Other than places, spatial relationships from the same discourse provide contextual knowledge as well. For example, relative direction relationships can be used to infer the reference direction used by the descriptor, especially when using local landmarks as *relata*. The inferred reference directions can help with interpreting other relative direction relationships in the discourse, and thus be used to locate places as *locata* of these relationships.

The order of appearance of places and relationships in a place description should also be preserved. For example, descriptors often switch the level of spatial granularity monotonically e.g., changing from city-level to district-level [47]. Such changes in context cannot be detected without recording the order of appearance of place references and spatial relationships. Similarly, reference directions can also change within a description, for example at turns, and affect the interpretation of subsequent relations.

Storing sequential order also helps linking different place references that are referring to the same place. Definite references such as ‘the building’, which refers to a building described previously in the discourse, can be ambiguous without sequential appearance information, if there were multiple buildings mentioned in the discourse.

Information about places and relationships from the same discourse, as well as their sequential order of appearance is not modelled in the original place graph model. Triplets from different descriptions are merged without any indexing mechanisms for future separation. The two types of information can be obtained directly without requiring an additional parser; the challenge is how to modify the place graph model in order to store this knowledge.

3.1.3. Reference frame and direction

The original place graph does not capture spatial reference frame and reference direction information [57,58]. Anchoring relative direction relations is, thus, problematic, as it is unknown which directions are being referred to. It is also difficult to perform qualitative spatial reasoning (QSR) [59,60] or to interpret seemingly contradicting direction relations, as in the example <the Arts Faculty Building, *left*, the Old Quad> and <the Arts Faculty Building, *right*, the Old Quad>, without knowing the reference directions used in both situations.

Reference frames in natural language have been classified in the literature [58]. In this research, a relative direction reference frame is defined to be either intrinsic or relative. For example, the expression ‘the café is in front of the library’ is likely to use the intrinsic reference frame of the library, which has a front, while ‘you will find the library to the left side of the lawn’ is likely to use the relative reference frame of the walking direction, since the lawn has no front (or left). A parser for identifying reference frames and heading directions is not yet available. Nevertheless, we will demonstrate how these two types of information can be modeled in an extended place graph, as well as how they can be leveraged in application scenarios.

3.1.4. Non-binary relationships

Non-binary spatial relationships, e.g., *across*, *between*, *around*, and *among*, involve more than two places thus cannot be represented by the aforementioned triplet structure. Vasardani *et al.* suggested for the original place graph that ternary relations can be modeled by two edges linking two relatum nodes to the same locatum node [1]. However, these edges are not indexed hence can become ambiguous when several of these relationships are known for one place. Furthermore, several of these relationships can have more than three places involved. For non-binary relationships, the task in this research is how to properly model them in order to preserve the original semantics and to allow future tracking.

3.1.5. Number of occurrences of place references and spatial relationships

The original place graph does not store the number of times each place reference is used to refer to a place, and thus, the information of which references are more frequently used for a place is lost. Storing the number of occurrences for place references can distinguish between common (popular) names and less-frequently used ones. It also enables analyzing which references are more often used in certain conversational contexts, description themes, or by certain people.

The number of occurrences of each relation being used between two places is not recorded in the original model either, as only one instance for each relation can exist between any two nodes. As a result, if two contradicting relations *north of* and *south of* between two places have both been stored in a place graph, it is impossible to determine which one is more likely to be the true, according to frequencies. By preserving the number of occurrences for each relation, the one that occurs more often can be regarded as a better-agreed upon assertion and, thus, more likely to be true.

3.1.6. Conceptualization of places

According to Lynch's classification of elements of the city [61], a place from an urban environment can be conceptualized as a node (a strategic spot that is accessible), a path (a channel that affords movement of the observer), a district (an accessible and identifiable area), a landmark (an inaccessible place typically for spatial referencing), or an edge (an inaccessible boundary), as a 0D, 1D, or 2D object. The classification has been adopted in geographic information science, such as for describing the functional spatial structure of urban environments using graphs [62].

The sense of place emerges as it is functionally different from its surrounding environment and, thus, becomes distinguishable. The functional difference between places is sometimes revealed by place conceptualization in descriptions, and such difference is context-dependent. The same place can be conceptualized differently in different description contexts or even within the same description [19], depending on what information the descriptor wishes to convey. For example, a district can be regarded as a 2D container for describing places within it, or being regarded as a 0D landmark for locating other nearby places, either from the same granularity level or not.

We argue that capturing the conceptualization of places in descriptions allows for better interpretation of the information communicated. For example, the same description 'the place is to the north of the campus' can either be interpreted as an external cardinal direction relationship (mapped as *north* and *disjoint*) or an internal one (mapped as *north* and *inside*), depending on the conceptualization of the relatum, i.e., if the descriptor is regarding the campus as a landmark and describing places nearby, or as a container and describing places within. In the examples, the conceptualization of a place can be regarded as a variable that affects the mapping of vernacular spatial relationship expressions to formal relations. Without capturing place conceptualizations, the mapping process becomes either risky or unrestrictive.

3.1.7. Route and accessibility

Some place descriptions can take the perspective of a route description. Route descriptions are often associated with reference directions and accessibility information for navigation purposes. For a triplet in a route description, the accessibility from the relatum to the locatum is usually implied. Accessibility also determines whether the triplet belongs to a part of a route or not. For

example, Moncla *et al.* use motion expressions to distinguish places that are only seen and places that can actually be reached from hiking descriptions, in order to reconstruct itineraries [63]. In the GUM ontology [48], a relationship indicating accessibility is classified as a *GeneralizedRoute*, and is distinguished from a *GeneralizedLocation* which does not belong to a route.

Tracking whether places and relationships originated from a route description enables querying of path knowledge for purposes such as navigation support. Moreover, as the number of occurrence of relationships is also preserved in an extended place graph, it is also possible to identify prominent routes that are described more often by people.

3.1.8. Description context and source context

Some relations can vary with context. For example, Yao and Thill identified several contextual variables that determine the choice of qualitative distance relationships, e.g., the current type of activity, and the available mode of transportation [64]. Also, the places referred to in a place description are depending on the purpose of the description. Therefore, information of the theme of a description is useful for place analysis. For example, Kim *et al.* observed the differences of place occurrences in descriptions of four themes: environment, business, travel, and other [10]. In their implementation, places from their place graphs have to be manually re-link to their original descriptions, as the correspondences between places and occurrences in descriptions are not kept when the graphs were constructed. Identifying thematic topics of textual documents can be done using existing techniques such as topic modelling [37,38].

From a database perspective, the metadata of a place description, e.g., its source and time-stamp, should also be preserved. Such knowledge can be useful when determining the reliability or time validity of the extracted knowledge. Thus, original speakers and recipients form a facet of context. Individuals may describe the same environment differently, in terms of the selection of place and place semantics, spatial relations, reference frame, and conceptualization. The intention of giving a description to a recipient (or recipients) may also influence how a description will be organized [52]. Human-level factors such as age, gender, ethnicity, and degree of familiarity with the environment have also been identified as being influential on the meaning of the spatial relationships communicated [64]. In addition, the affordance of a place also varies among individuals. For instance, a supermarket may afford for some people shopping, and for others (or at other times) work.

Linking descriptions with people allows richer types of queries and analysis to be performed on an extended place graph, e.g., what places are more frequently mentioned by whom. As another example, an extended place graph database supporting an autonomous vehicle can be used to establish links between passengers' accounts and the place each one of them calls *my home*. Such human-related information is often unavailable or limited, for example when descriptions are collected from online documents.

3.2. The extended place graph database model

The extended place graph database model is illustrated by a UML diagram shown in Figure 5. The model preserves all the additional information specified so far and is designed to support efficient querying through graph traversal. Each class in the diagram represents a node type in an extended place graph, and each relationship represents an edge type. All node types and some edge types are associated with properties. Values of some of the properties will be set to *null* if the corresponding information is not (or not yet) available. For example, properties such as *footprint* of a place node cannot be obtained directly from place descriptions, and must be derived using georeferencing techniques [50].

An *n*-plet is an extension of a triplet, and each place reference that occurs in a place description is regarded as being embedded in an *n*-plet. An *n*-plet is often a triplet representing a binary relationship; however, it can also represent a non-binary relationship, e.g., *between*, *around* and *across*, having multiple locata and relata based on the sequential order of appearance in the description. An *n*-plet can also consist of only one locatum without any relatum, as a place reference may not be embedded in any locative expressions in a description, e.g., 'Melbourne is a

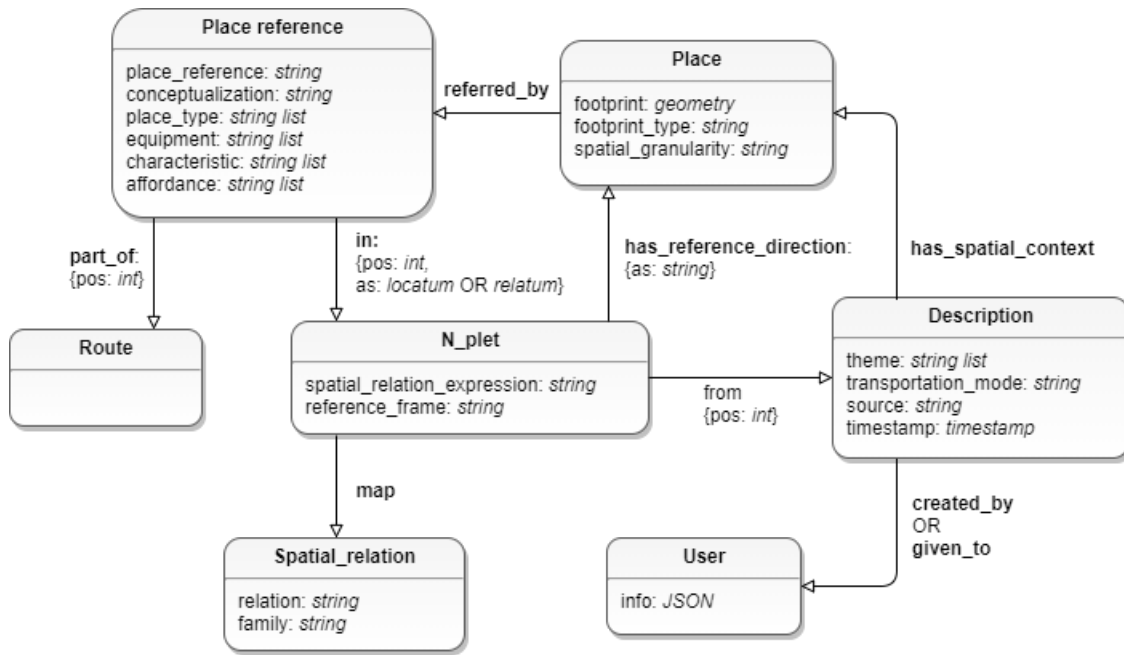


Figure 5. UML diagram illustrating the extended place graph database model, with seven types of classes (nodes) and nine types of relationships (edges).

populous city’. Thus, an n -plet must have at least one locatum, and any non-negative number of relata. In the remaining part of the section, each type of node, edge, and the associated properties are discussed.

3.2.1. Place reference node

A *place_reference* node represents a reference to a place from an n -plet in a description, either as a locatum or a relatum. Each place reference node must have one and only *referred_by* incoming edge from a *place* node. Between place nodes and place reference nodes are $n:1$ relationships, i.e., a place may be referred to by one or more different place references, while the same place reference may be used to describe different places in different contexts (but modelled by distinct place reference nodes). For example, two references ‘Flinders Street Railway Station’ (an official place name) and ‘the train station’ (a non-gazetteered reference) come from conversational contexts where they refer to the same place (Flinders Street Railway Station). In a different context, the reference ‘the train station’ may refer to another train station.

A place reference node, when created, is by default linked to a new place node instance through a *referred_by* edge. A merging algorithm [9] can then modify the correspondence by removing the newly created edge and place node instance, and establishing another *referred_by* edge between the place reference node and a pre-existing place node, if it is determined so.

Since place references are embedded in n -plets extracted from place descriptions, each place reference node has one and only outgoing edge *in* to an n -plet *N_plet* node. An *in* edge has two mandatory properties: *pos* and *as*. The value of *as* can either be *locatum* or *relatum*, representing whether the place reference is corresponding to the locatum or the relatum of the n -plet. The property *pos* is a positive integer denoting the index of the occurrence of the place reference, as it is possible that an n -plet has multiple locata or relata. For a triplet, the value of *pos* is 1 for either of the two place reference nodes it links to.

Thus, a place reference node is defined by Axiom 1 below:

$$\text{Place_reference} \sqsubseteq \exists \text{referred_by}^-. \text{Place} \sqcap \exists \text{in}. \text{N_plet} \quad (1)$$

A place reference node has six properties: *place_reference*, *conceptualization*, *place_type*, *equipment*, *characteristic*, and *affordance*. Among the properties only *place_reference* is

mandatory. The value of *conceptualization* is one of the categories based on Lynch's classification: *node*, *path*, *district*, *landmark*, or *edge*. Values for the remaining four properties are unrestricted, and some examples are given in Section 3.1.1. The data type of these four properties is string list, as multiple values of each of these properties can be described. These properties are not stored under place nodes in order to preserve the context of where and by whom these values are given.

3.2.2. N-plet node

An *n*-plet node is defined by Axiom 2 below. Other than *in* as already explained, each *n*-plet node has one and only edge *from* as an outgoing edge, denoting the description from which this *n*-plet is extracted. The edge *from* has the same property *pos* as *in*, showing the sequential order of appearance of an *n*-plet in the description. An *n*-plet node can have one or more *in* and *map* edges, depending on the number of locata and relata, and the number of mapped formal spatial relations mapped respectively.

$$N_plet \sqsubseteq \exists in^-.Place_reference \sqcap \exists from.Description \sqcap \exists map.Spatial_relation \quad (2)$$

An *n*-plet node has two properties: *spatial_relation_expression* and *reference_frame*. The first one stores the original spatial relationship expression used for the *n*-plet in the description. In an original place graph, such expressions are formalized by a controlled vocabulary before graph construction [51], yet it is quite often that the same spatial relationship expression can be mapped to different formal relations depending on the context. Therefore, in an extended graph, the original spatial relationship expressions are kept, and the mapped relationships will be stored separately as *spatial_relation* nodes linked by outgoing edges *map* from *n*-plet nodes.

If the spatial relation expression of an *n*-plet is mapped to a relative direction relation, the value for the property *reference_frame* can either be *intrinsic*, *relative*, or *null* (undetermined). The *intrinsic* value means the relative direction is based on the intrinsic direction of the relatum (e.g., 'in front of the building'), while the *relative* value means a non-intrinsic reference direction is adopted. If the value *relative* is used, the *n*-plet node will have an additional outgoing edge *has_reference_direction* to a place node referred in the discourse, anchoring the reference direction used for the *n*-plet. A *has_reference_direction* edge has a property *as*, and the value is one from {*front*, *back*, *left*, *right*, *left front*, *right front*, *left back*, *right back*}. An example is given in Figure 6, with description:

"... coming from the **Main South Entry**, the **Baillieu Library** will be **on the left hand side** of the **South Lawn** ..."

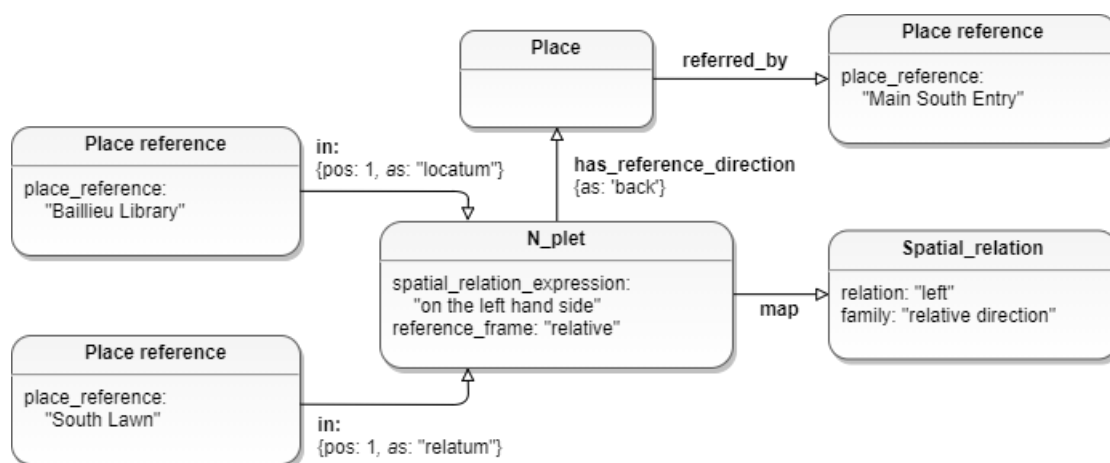


Figure 6. An example of modelling a relative direction relationship using the extended place graph.

3.2.3. Place node

Each *place* node represents a place. In an extended place graph, a place is identified from one or more place descriptions by place references embedded in *n*-plets. A place node does not have any place references stored; however, all the references used for referring to it (as well as the number of occurrence for each reference) can be obtained easily from all the place reference nodes it is connected to through outgoing *referred_by* edges. A place node is defined by Axiom 3:

$$\text{Place} \sqsubseteq \exists \text{referred_by}.\text{Place_reference} \tag{3}$$

A place node has three derived properties. The value of *footprint* represents the location of the place, and can be either a point, a polyline, a polygon, or an *approximate location region* (ALR) [50]. An ALR is a region derived using spatial relation search space models (including formal and probabilistic models) for georeferencing places without gazetteered references. The value of property *type* denotes the data type of the footprint, e.g., polygon. The property *spatial_granularity* is a classification of the spatial granularity of the place based on the categories found in [65]: {*furniture, room, building, street, district, city, country*}.

3.2.4. Route node

Places referred to as part of a route are grouped by linking their corresponding place nodes to a route node through *part_of* edges. The property *pos* of a *part_of* edge records the position of the place reference in the route by sequential order of appearance, and the value is a positive integer.

3.2.5. Spatial relation node

Each *spatial_relation* node represents a formal spatial relation. Unlike a value of the property *spatial_relation_expression* stored in an *n*-plet node, which can be expressed in flexible ways, formal relations are from a controlled vocabulary. Binary formal relations from four families are considered, as listed in Table 1. The vocabulary of non-binary relations is not restricted, since non-binary relations have not yet been well defined in literature. Accordingly, the value of property *family* is from one of the five spatial relation families: {*qualitative distance, cardinal direction, relative direction, topological, non-binary*}. The property *relation* stores the name of the relation.

Mapping between spatial relation expressions and formal relations is a *m:n* relationship. A *spatial_relation_expression* value can be mapped to one or more formal relations from single or multiple families, and different spatial relation expressions could be mapped to the same formal relation. The mapping process is context-dependent. For example, a spatial relation expression ‘north’ can be mapped to either *north, disjoint* (external north) or *north, inside* (internal north), depending on the original expression and place conceptualization. Compared to an original place graph, the extended model supports more flexible and context-aware reasoning of spatial relations.

Table 1. Formal binary spatial relations considered in this research.

Spatial relation family	Spatial relation type
Cardinal direction	north, south, east, west, northeast, southeast, northwest, southwest
Qualitative distance	near
Relative direction	front, back, left, right, left front, right front, left back, right back
Topological	inside, covered by, overlap, meet, disjoint, cover, contain, equal

3.2.6. Description node

A *description* node represents a description document as a single discourse. It is used to store the global-level context variables of the descriptions from which *n*-plets were extracted. The four properties of a description node, i.e., *theme, transportation_mode, source, and timestamp*, have already been explained in Section 3.1.8.

A description node has at least one instance of an incoming edge *from* from an *n*-plet node. The property *pos* of *from* is the position of the linked *n*-plet in the description by appearance, and the value is a positive integer. The property *spatial_context* is a derived one, representing the geographic extent a description is embedded in, using the approach developed in [50]. For example, if the extracted places are landmarks in the Melbourne CBD as a suburb, the context of the original description is likely to be about Melbourne CBD. Finally, a description node can also have outgoing edges *created_by* and *given_to* to a user node, if such information is available.

3.2.7. User node

A user node either represents a descriptor (connected by a *created_by* edge from a description node) or a recipient (connected by a *given_to* edge). The same user node can be connected to multiple description edges by either roles. The property *info* of a person node is not restricted in this research, as what information is useful for the application of an extended place graph is domain- and task-dependent. Examples have been given in Section 3.1.8. The property value is defined in the format of JSON as key-value pairs.

3.3. Summary

The purpose of the extended place graph model is to capture information ignored in the original place graph, in order to overcome the graph's limitations. The extended place graph captures the five core concepts of spatial information proposed by Kuhn [66]: *location*, *field*, *object*, *network*, and *event*. The graph stores the *location* of a place, and the probabilistic ALR derived using the approach in [50] represents the approximate location of the place as a *field*, based on its relationships to other places from the discourse. Different places are modeled as node *objects* in a place graph, characterized by information including place references and semantics. An extended place graph also forms a *network* by representing the links not only between places, but also between places and descriptions, as well as between places and people. Such links can be strengthened given additional descriptions, as their times of co-occurrence are captured as well. Finally, an *event* involves aspects of people, time, location, and activity, and these aspects are covered by node properties *info* (of user nodes), *timestamp* (of description nodes), *footprint* (of place nodes), and *affordance* (of place reference nodes) respectively.

4. Implementation and Experiments

This section describes the implementation of the extended place graph model, as well as three experiments to demonstrate how an extended place graph outperforms the corresponding original one. We do not use all the identified information in Section 3.1 in the experiments, since the goal is not to provide a comprehensive evaluation, but rather to demonstrate the superiority of the extended model. We also discuss additional challenges and insights.

4.1. Data overview and construction of the test place graph database

The description dataset used in this research are place descriptions submitted by 42 graduate students about the University of Melbourne campus environment. The data set consists of over 9000 words, 726 *n*-plets, and a total of 241 distinct places being mentioned. A place can be referred to by multiple place references in different descriptions, or even within the same description. Some places are not referred to by any gazetted names due to two reasons: the place itself is not captured in the gazetteer, possibly due to granularity (e.g., 'the small courtyard', 'the dean's office'); or the place is only referred to by synonyms or other vernacular names as commonly-used references (e.g., 'the mathematics department building'). A part of a description is shown below:

"... If you go into the Old Quad, you will reach a square courtyard and at the back of the courtyard. You can either turn left to go to the Arts Faculty Building, or turn right into the John Medley Building and Wilson Hall. Raymond Priestly Building is the open aired ground area which is in front of Wilson Hall that is adjacent to it. Towards North, which is when you turn left when exiting the Old Quad, you will see Union House where

there are shops selling foods. If you continue walk along the road on the right side where you're facing Union House, you can see the Beaurepaire and Swimming Pool. There will also be a sport tracks and university oval behind it ..."

A parser was used in previous studies to extract triplets from the dataset [8]. The extracted triplets are stored in a csv file with three columns: locatum place reference, spatial relation, and relatum place reference. However, such a csv file does not preserve the necessary information for constructing an extended place graph. Therefore, we create a JSON file in order to capture additional information. For the following experiments, the information of reference direction is required as well. Due to lack of a parser capable of extracting this information, a graduate student was asked for manual annotations to obtain the information. To minimize the influence of pre-existing local knowledge on the annotation process, all place references were replaced by five place types: *building*, *spot* (any place finer than a building), *area*, *alley*, and *street*. For example, the first building name that occurs in a description is anonymized as *Building_1*. The task of the student was to assign each relative direction relationship with three property values: reference frame (*intrinsic* or *relative*), the anonymized reference of the place indicating the reference direction (e.g., 'Main South Entry' in Figure 6), and the reference direction (value of property as for edge *has_reference_direction*). The structure of the final input JSON file for creating the test place graph database is shown below:

```
{
  "descriptions": [
    ...
    {
      "did": 2,
      "nplets": [
        {
          "nid": 4,
          "locatum_reference": "Baillieu Library",
          "spatial_relation_expression": "on the left hand side",
          "relatum_reference": "South Lawn",
          "reference_frame": "relative",
          "reference_direction": ["Main South Entry", "back"],
          "relation_map": ["left"]
        },
        { ... },
        { ... }
      ]
    },
    ...
  ]
}
```

A place graph database management system interface has been implemented using Neo4j and Python, as shown in Figure 7. The system is able to perform tasks including description parsing, graph creation from JSON, graph visualization, georeferencing, qualitative reasoning, querying, and mapping. Some functions are necessary for the following experiments and will be explained in detail in the remaining part of this section.

4.2. Experiment I: locating places without gazetteered references

An approach has been developed previously to georeference places in an original place graph [50]. The approach first identifies and disambiguates places with gazetteered, i.e., well known references, called *anchor places*, based on gazetteer look-up and a clustering-based disambiguation approach. Then, the anchor places are used to derive the spatial context of the graph. Next, an ALR is derived for each remaining place node by intersecting the spatial context, as well as the search spaces for the spatial relationships between this place and the anchor places. Search spaces can either be based on formal or probabilistic models. An illustration is provided in Figure 8 based on formal search space models. The location of place *b*: *Federation Square* is represented by the shaded region, which is derived by intersecting the search spaces of the three relationships: *east of a*, *south of c*, and *near c*.

In this experiment, three ALR refinements are possible, leveraging two of the newly captured information: identifying places from the same discourse, and reference direction. The first refinement is by separating spatial contexts for individual descriptions. The original place graph merged from different descriptions has only one spatial context. As shown in Figure 9 (left), for places from different descriptions, only one spatial context will be derived. Consequently, the ALRs generated for places in the first description will be inappropriately large. However, separate

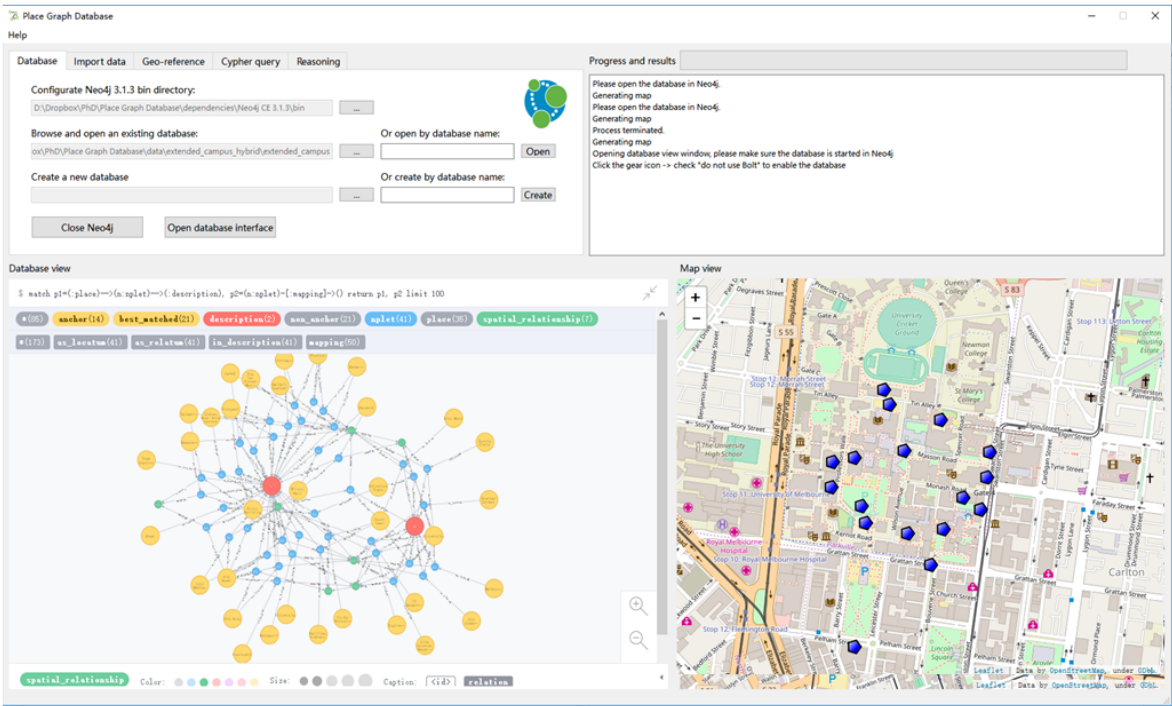


Figure 7. The implemented extended place graph database management system interface, with an example visualization of part of the test extended place graph.

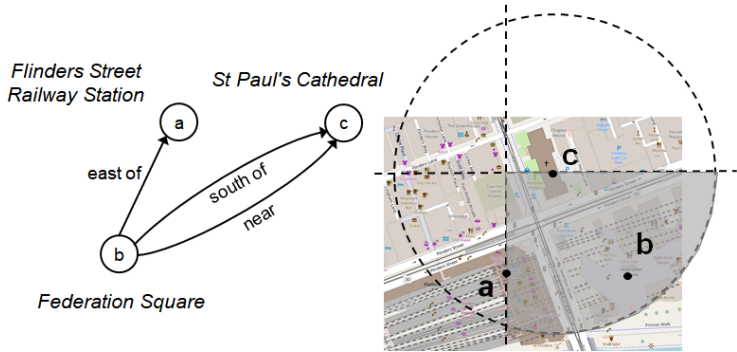


Figure 8. An example of deriving the ALR for place *b* through intersection of search spaces.

spatial contexts regarding each description can now be derived (Figure 9, right), since the link between descriptions and place references is preserved in the extended place graph.

The second refinement is by anchoring relative direction relations using the newly captured reference direction information. Search spaces for relative direction relations can only be defined as buffered regions similar to *near* information, as shown in Figure 10 (left). The reference direction of a relative direction relation can only be anchored, if the locations of the relatum and the place indicating reference direction (i.e., the place linked by the *has_reference_direction* edge) are available. The proposed refined search is illustrated in Figure 10 (right), with *front* indicating the reference direction, and the shaded regions representing search spaces.

Third, the interpretation of qualitative distance relationships, such as *near*, can be contextualized by considering other relevant places mentioned in the discourse, based on the theory of contrast sets [19]. An example is given in Section 3.1.2. In this experiment, the contrast set of a relatum is defined as the set of places that are of the same granularity level as the relatum, and have been used as relata in the same description. The underlying assumption is that when a descriptor says ‘A is near B’, it is often implied that A is closer to B than any of the other places

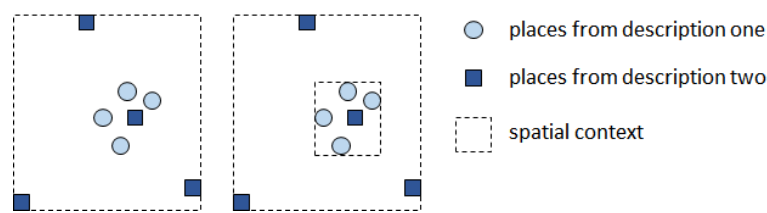


Figure 9. The spatial context of a merged, original place graph (left), and separated spatial contexts of an extended place graph (right).

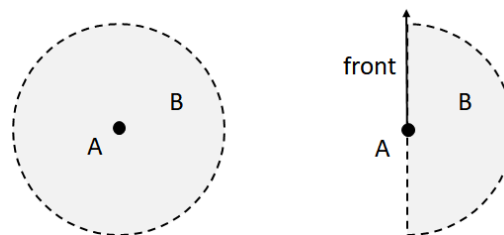


Figure 10. Search space of place *B* for relationship $\langle B, \text{right of}, A \rangle$ without a reference direction (left) compared to with anchored reference direction information (right).

used as relata in the discourse, and are of the same granularity as *B*. Similarly, Vasardani *et al.* provided a first interpretation of the preposition *at* by contrast sets using their Voronoi diagrams [40]. In this research a similar approach is adopted, as illustrated in Figure 11. The shaded region indicates the search space derived based on a contrast set, and which is intersected with other search spaces to derive the ALR of the locatum in the georeferencing process.

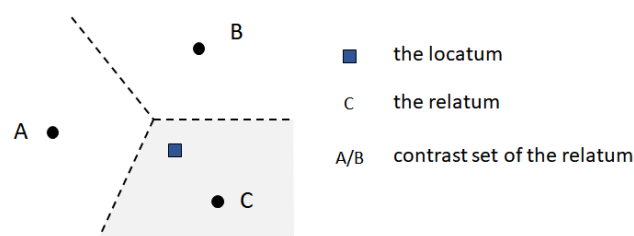


Figure 11. The search space of a qualitative distance relationship with contrast set information, represented by the shaded region.

4.3. Experiment II: relational consistency reasoning using reference direction information

This experiment aims at leveraging reference direction information to determine the relational consistency of relative and cardinal direction relations stored in an extended place graph, e.g., determining whether the two relationships $\langle \text{the Arts Faculty Building, left, the Old Quad} \rangle$ and $\langle \text{the Arts Faculty Building, right, the Old Quad} \rangle$ are contradicting, or not. This experiment is not about checking global, but rather local consistency. Relational composition is also not considered.

Two complementary reasoning approaches are suggested in this work. One is based on search space intersection, and the other relies on reference direction transformation. An illustration of the first approach is given in Figure 12, where the shaded regions are search spaces as explained in the previous experiment. The approach first finds all directional relationships between two places in an extended place graph, and derives their search spaces. If any pair of these search spaces does not intersect, an inconsistency is flagged.

However, deriving such search spaces requires the location information of the relata as well as the places indicating the reference directions, which may not always be available. Therefore, a complementary qualitative reasoning approach is proposed using reference direction translation

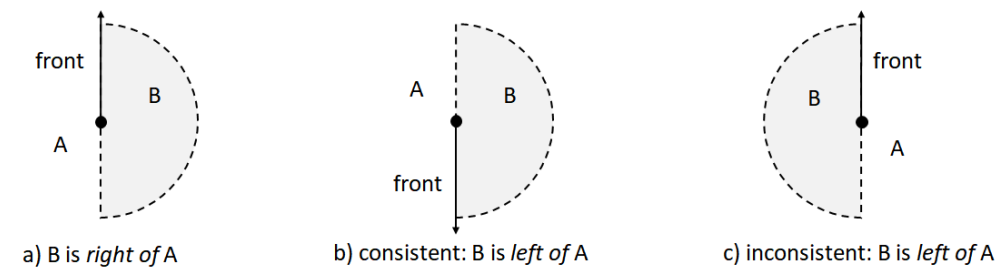


Figure 12. Determining consistency of directional relationships between a locatum *B* and a relatum *A* by search spaces.

rules. An illustration is provided in Figure 13. Given the knowledge in (a), consistent relationships can be inferred as shown in (b) and (c), through translations of the reference direction. Thus, given another relationship ‘*A* is *right* of *B* with *C* as *back*’ (meaning with *C* indicating the back direction of the reference direction)’, it can be identified as inconsistent with (a), (b), or (c). The drawback of this approach is that it is only applicable to scenarios with up to three places, while the previous approach does not have this limitation. The full reasoning procedure of this experiment is described in Algorithm 1.

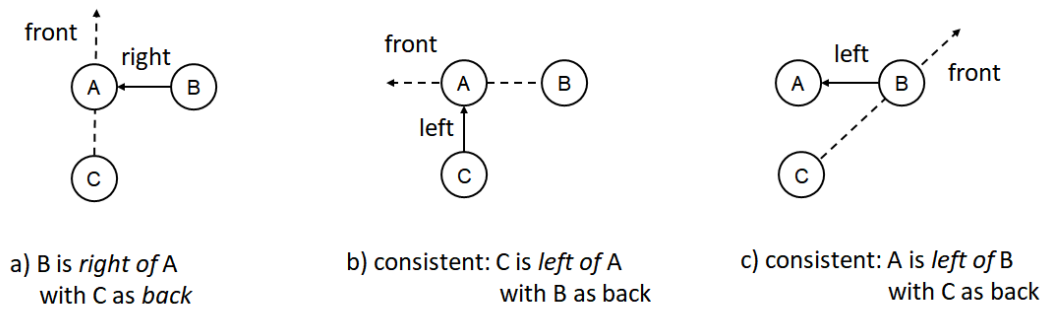


Figure 13. Consistency reasoning through reference direction translation.

4.4. Experiment III: spatial knowledge querying

All the queries to the original place graph can still be performed in the extended place graph, e.g., ‘find the relationship between two places’, or ‘find places that have relationships to a particular place.’ An extended place graph additionally supports queries complex queries, such as finding in which descriptions a particular place occurred, as well as other places mentioned in the same descriptions ranked by their co-occurrence frequency.

Cypher queries are used in an extended place graph. For example, the corresponding Cypher query for the NL query ‘find computer labs that are inside the University of Melbourne’ is shown below. A Cypher query is a graph traversal algorithm that attempts to find nodes or edges that match certain label (place and edge type) and properties values. In this research, results are returned simply by criteria string matching.

Three query examples that cannot be answered with the original place graph are selected:

- Find the most frequently referred to relatum (landmarks).
- Find places that are most frequently linked to a specific place by spatial relations (place relevance by co-occurrence).
- Find the most frequent paths of length three, consisting of only directional relationships, i.e., *place A*-*relation a*->*place B*-*relation b*->*place C* (prominent routes)

```
MATCH (p:place)->(a:place_reference)-[:in {as: 'locatum'}]->(b:n_plet),
```

Algorithm 1 Consistency reasoning of directional relationships.**Input:** *place_graph***Output:** *inconsistent_pairs*

```

1: inconsistent_pairs :=  $\emptyset$ 
2: for relationship in getDirectionalRelations(place_graph) do
3:   search_space = computeSearchSpace(relationship)
4:   locatum, relatum, reference_direction_place = getPlaces(relationship)
5:
6:   others = getDirectionalRelationsBetween(locatum, relatum)
7:   for new_relationship in others do
8:     new_search_space = computeSearchSpace(new_relationship)
9:     if intersect(search_space, new_search_space) == False then
10:      inconsistent_pairs  $\leftarrow$  (relationship, new_relationship)
11:    end if
12:  end for
13:
14:  others = getDirectionalRelationsWith(locatum, relatum, reference_direction_place)
15:  for new_relationship in others do
16:    if new_relationship not in transformation(relationship) then
17:      inconsistent_pairs  $\leftarrow$  (relationship, new_relationship)
18:    end if
19:  end for
20: end for
21: inconsistent_pairs.removeDuplications()
22: return inconsistent_pairs

```

```

(c:place_reference)-[:in {as: 'relatum'}]->(b:n_plet)-->(d:spatial_relation)
WHERE a.place_type = 'computer lab' and
c.place_reference = 'the University of Melbourne' and
d.relation = 'inside'
RETURN p

```

5. Results and Discussion

This section discusses the results of the three experiments. In Experiment I, places without gazetteered references have been georeferenced using five methods, namely baseline (without using any refinement approaches), SC (applying only the spatial context separation refinement), RF (applying only the reference direction refinement), CS (applying only the contrast set refinement), and Hybrid (all of the three refinements are applied). Refinement methods can only reduce the size of the ALR derived for a place. Thus, for each of the latter four refined georeferencing methods, four results are possible when georeferencing a place:

- The size of the ALR is reduced compared to the one from the baseline, but both ALRs capture the ground-truth location of the place (Case 1).
- There is no change in the ALR's size (Case 2).
- The ground-truth location is not captured in the either ALR (Case 3).
- The ground-truth location is captured by the ALR of the baseline method, but not in the reduced-size ALR (Case 4).

Figure 14 shows the percentages of places that belong to each of the four cases, grouped by the four georeferencing methods. Places from Case 1 are regarded as better-georeferenced, while places from Case 2 and 3 are considered as equally georeferenced, and those from Case 3 are regarded as worse-georeferenced, when compared to the baseline. Among the first three methods, SC has the largest proportion of better georeferenced places, while for the RF and CS methods the percentages are much lower. In order to get refined ALRs for the latter two methods, relative direction relationships to some anchor places with reference direction information (for the RF method), or anchor places in the discourse as members of a contrast set (for the CS method) must be available. Since that is not always the case, only part of the places to be georeferenced can benefit from these two refinement methods.

When applying RF and CS methods, some places are worse-georeferenced (Case 4). For the RF method, this is because some relative direction relation information is incorrect, either due to

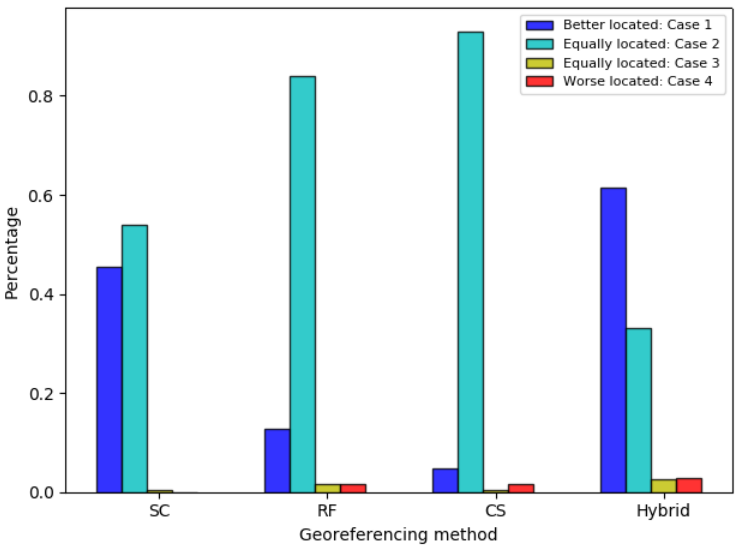


Figure 14. Percentages of places from different ALR refinement situations compared to baseline.

mistakes made by descriptors, or the imperfection of the reference direction annotation procedure. For the CS method, the worse-georeferenced cases are because some ALRs are over-refined and, thus, not capturing the ground-truth locations of places anymore.

The hybrid method has overall the largest improvement in terms of the proportion of better-georeferenced places. This is expected as the method requires only information that can be used in any one of the three previous methods in order to make refinements. One drawback is that the hybrid method also has the largest worse-georeferenced place numbers due to error propagation. It is a trade-off problem between sizes of ALRs (the smaller the better as being more constraining) and having ALRs capturing the ground-truth location of places. A measure of the refinement in terms reduced ALRs is depicted in Figure 15, which shows in percent the ALR remaining size after refinement compared to the baseline, for individual places. A value of 0.6, for instance, means the refined ALR is 60% of the size it was in the baseline method. Only places with the available required information for refinement are included in the figure. For example, for the RF method, only places with relative direction relationships and reference direction information to some anchor places are included. The hybrid method results in the most size reduced ALRs for all places, which is also expected, given the method uses all the refining information available, combining the restrictions of the other three methods.

Table 2. Results for the three queries from the third experiment, with the second query using the place *Alice Hoy* as an example.

Rank	Most frequent relata	Places most frequently co-occurring with <i>Alice Hoy</i>	Most frequent length-3 paths
1	University Of Melbourne	Monash Road	<Old Arts, <i>right</i> , Baillieu Library, <i>left</i> , South Lawn>
2	Union Building	Entrance	<Baillieu Library, <i>left</i> , South Lawn, <i>front</i> , John Medley>
3	Grattan Street	University of Melbourne	<Royal Parade, <i>left</i> , Baillieu Library, <i>left</i> , South Lawn>
4	South Lawn	Wilson Hall	<Medical Building, <i>left</i> , Baillieu Library, <i>left</i> , South Lawn>
5	Swanston Street	Peter Hall Building	<Baillieu Library, <i>left</i> , South Lawn, <i>left</i> , Wilson Hall>

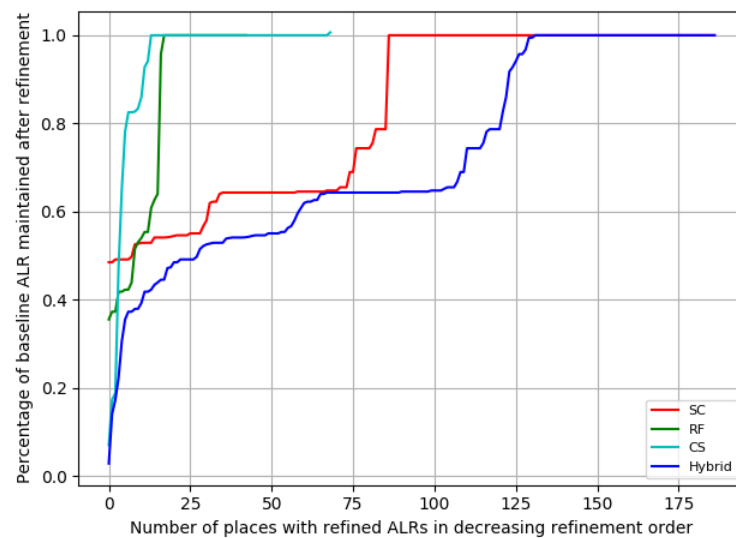


Figure 15. Refined ALR sizes as percentages of the original (baseline) ALR size.

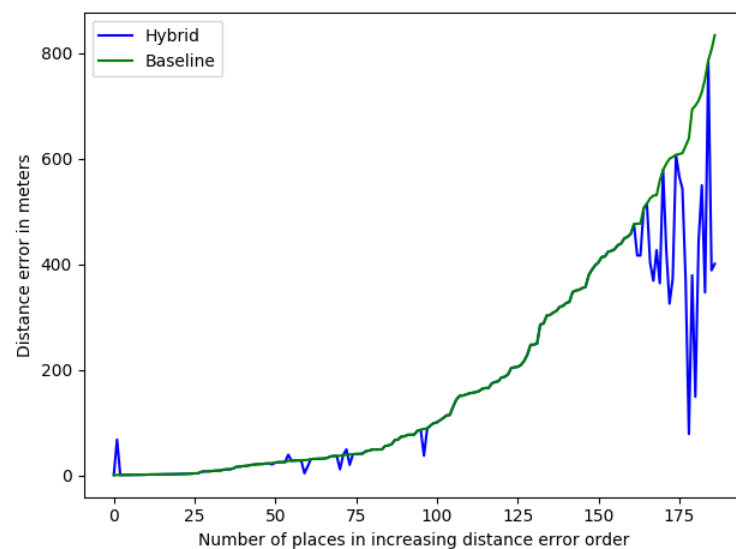


Figure 16. Distance errors between ground-truth and matched gazetteer locations for the baseline and hybrid methods.

Places to be georeferenced are matched to gazetteer entries that fall within their derived ALR, based on the method described in [50] considering string, semantic, and spatial similarity. For example, if there is a gazetteer entry named ‘University Square’ within the ALR derived for the place reference ‘the large square’, the two place references are likely to be matched. The distance errors between ground-truth and matched gazetteer locations for individual places are shown in Figure 16 for the baseline and the Hybrid method, sorted by error size in the baseline. Large distance errors in the baseline seem to be more likely to be reduced by the hybrid refinement method. A possible explanation is that large distance errors usually correspond to large, less restricting ALRs, and in such situations a refinement is more effective. On the other hand, if an ALR derived in the baseline is already constraining, further refinement might have no, or even negative effects. For example, the peak on the left side of the axis represents a place that was correctly linked to its corresponding gazetteer entry in the baseline. In contrast, the refined ALR

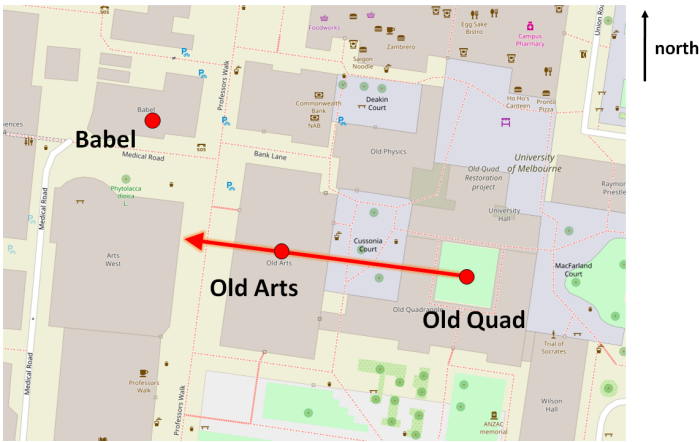


Figure 17. The locations of the three places mentioned in the descriptions above, with a red arrow indicating the walking direction.

leaves out the ground-truth location, and causes the place to be miss-matched, resulting in an increased distance error.

In Experiment II, directional relations in the extended place graph are checked for relational inconsistency using Algorithm 1. Among the 43 directional relations (out of 726 in total) stored in n -plets, four were identified as inconsistent with other relationships stored in the database. As an example, consider the following two descriptions of the environment in Figure 17:

“... You’re now in the **Old Quad** ... Pass through the **Old Arts** building and immediately look to your left – the tall building is the **Babel** building that, somewhat ironically, houses the languages and linguistics departments ...”

“... From the **Old Quad**, you can go through the **Old Arts** building, and then turn right and walk until you come to a building called the **Babel** building (a 1970s yellow brick monolith) ...”

The two relationships between the Babel building and the Old Arts Building from the two different descriptions are denoted here as ‘Babel *left of* Old Arts with Old Quad as *back*’ and ‘Babel *right of* Old Arts with Old Quad as *back*’, respectively. The first description is not true as can be verified from the map in Figure 17. The algorithm successfully identified the two relationships as being inconsistent. Note that this reasoning mechanism only flags inconsistent relationship pairs, instead of deciding which one is true.

Table 2 shows results generated for the three selected queries, in Experiment III. The second column ranks the most frequent relata in the test dataset, which can be regarded as local landmarks considering their prominence in the descriptions. In the original place graph, landmarks can only be identified by unique relationships, as the number of occurrences of neither place references, nor instances of the same spatial relation is preserved. For example, the relationship <Old Arts, *right*, Baillieu Library> was described more than ten times in the dataset. The original place graph stores the relationship once, which from a knowledge base perspective leads to loss of information about frequently mentioned, prominent relationships.

The third column shows places that are most frequently linked to *Alice Hoy*, a place from the test place graph being used here as an example. Again, such co-occurrence knowledge is not captured in the original place graph. The last column shows the top five most frequent length-3 paths, which are actually all from route descriptions.

6. Conclusions

Place descriptions occur in everyday communication as a way of conveying spatial information about place. This research proposes a graph-based approach for modeling and utilizing information

from place descriptions, including references to places, spatial relationships, place semantics, and various contextual knowledge. The model is a revision and an extension of the original place graph [1]. Place graphs are regarded as place-based knowledge bases that can be used in reasoning, georeferencing, landmark identification, and sketch-map drawing. However, place graphs constructed from triplets stripped off of context and other useful spatial or non-spatial information, have restricted applications. This research proposes an extended model in order to overcome those limitations. We implement the model in a graph database and demonstrate its superiority over the original place graph in three experiments.

We identified eight types of information that are embedded in place descriptions and not captured in the original place graph model: place semantics and characteristics, places and spatial relations from the same discourse, as well as their sequential order of appearance, reference frame, non-binary relationships, co-occurrence of place references and spatial relations, place conceptualization, route and accessibility, and description context and source context. We then designed a graph database schema for modeling these types of information, allowing convenient and efficient query through graph traversal. We implemented the schema using the Neo4j graph database for three experiments, and developed a management system interface capable of multiple place graph operations.

To demonstrate the superiority of the new model, three experiments were performed on an extended place graph created from descriptions about the University of Melbourne campus. The first experiment georeferences places without gazetted matches, using their spatial relationships to anchor places (gazetted places that are georeferenced first). We propose three refinement methods to reduce the areas of the approximate location regions generated for these places, when compared to the baseline method in [50]. More than 60% of the derived regions are reduced in size with a Hybrid approach that takes into consideration information from all three refinement methods, all based on information newly captured in the extended graph. The second experiment leverages reference direction information for identifying contradicting directional knowledge in an extended place graph. The results show that the approach can identify inconsistent relationships based on both qualitative and quantitative data. In the third experiment, we demonstrate how the extended place graph is capable for answering queries that cannot be answered by the original place graph.

The extensions of this paper were identified by our own needs for richer reasoning. We anticipate that additional future extensions are possible, since each step in the process is extensible. Thus, in future work additional information such as place semantics could be introduced, which enriches ALR refinement and defining search spaces for non-binary relationships. Also, query expansion (e.g., based on place semantic similarity and relational inference) and relevance-ranking (e.g., based on occurrence and description source) mechanisms can be further developed. Some of the identified information, however, requires new techniques or parsers to obtain automatically, which are not readily available.

The extended place graph model provides a complementary knowledge base to traditional GIS. A place-based information system can be used in applications requiring human spatial and non-spatial knowledge about places. The implemented extended place graph management system supports Cypher queries and returns tabular results with high interoperability. Finally, the knowledge modeled in extended place graphs can help with better understanding human descriptions as input to spatial services and, thus, support intuitive and smooth human-computer interaction.

References

1. Vasadani, M.; Timpf, S.; Winter, S.; Tomko, M. From Descriptions to Depictions: A Conceptual Framework. *Spatial Information Theory*; Tenbrink, T.; Stell, J.G.; Galton, A.; Wood, Z., Eds. Springer, 2013, Vol. 8116, *Lecture Notes in Computer Science*, pp. 299–319.
2. Elwood, S.; Goodchild, M.F.; Sui, D. Prospects for VGI research and the emerging fourth paradigm. In *Crowdsourcing Geographic Knowledge*; Springer, 2013; pp. 361–375.

3. Lieberman, M.D.; Samet, H.; Sankaranarayanan, J.; Sperling, J. STEWARD: Architecture of a spatio-textual search engine. 15th Annual ACM International Symposium on Advances in Geographic Information Systems; Samet, H.; Shahabi, C.; Schneider, M., Eds.; , 2007; pp. 186–193.
4. Moncla, L.; Renteria-Agualimpia, W.; Nogueras-Iso, J.; Gaio, M. Geocoding for texts with fine-grain toponyms: an experiment on a geoparsed hiking descriptions corpus. 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems; Huang, Y.; Schneider, M.; Gertz, M.; Krumm, J.; Sankaranarayanan, J., Eds., 2014, pp. 183–192.
5. DeLozier, G.; Baldridge, J.; London, L. Gazetteer-Independent Toponym Resolution Using Geographic Word Profiles. AAAI, 2015, pp. 2382–2388.
6. Kordjamshidi, P.; Van Otterlo, M.; Moens, M.F. Spatial role labeling: Towards extraction of spatial relations from natural language. *ACM Transactions on Speech and Language Processing* **2011**, *8*, 4.
7. Khan, A.; Vasardani, M.; Winter, S. Extracting Spatial Information From Place Descriptions. ACM SIGSPATIAL Workshop on Computational Models of Place 2013; Scheider, S.; Adams, B.; Janowicz, K.; Vasardani, M.; Winter, S., Eds., 2013, pp. 62–69.
8. Liu, F.; Vasardani, M.; Baldwin, T. Automatic Identification of Locative Expressions from Social Media Text: A Comparative Analysis. 4th International Workshop on Location and the Web; Ahlers, D.; Wilde, E.; Martins, B., Eds., 2014, pp. 9–16.
9. Kim, J.; Vasardani, M.; Winter, S. Similarity matching for integrating spatial information extracted from place descriptions. *International Journal of Geographical Information Science* **2016**, *1*, 1–25.
10. Kim, J.; Vasardani, M.; Winter, S. Landmark Extraction from Web-Harvested Place Descriptions. *KI-Künstliche Intelligenz* **2016**, *2*, 151–159.
11. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221.
12. Goodchild, M.F. Formalizing Place in Geographical Information Systems. Communities, Neighborhoods, and Health: Expanding the Boundaries of Place; Burton, L.M.; Kemp, S.P.; Leung, M.C.; Matthews, S.A.; Takeuchi, D.T., Eds.; Springer: New York, 2011; pp. 21–35.
13. Winter, S.; Baldwin, T.; Renz, J.; Tomko, M.; Kuhn, W. Place knowledge as a trans-disciplinary research challenge for Geographic Information Science. UCGIS Symposium; Mennis, J., Ed., 2016.
14. Egenhofer, M.J.; Mark, D.M. Naive geography. International Conference on Spatial Information Theory. Springer, 1995, pp. 1–15.
15. Goodchild, M.F. Formalizing place in geographic information systems. Communities, Neighborhoods, and Health. Springer, 2011, pp. 21–33.
16. Relph, E. *Place and placelessness*; Pion: London, 1976.
17. Winter, S.; Bennett, R.; Truelove, M.; Rajabifard, A.; Duckham, M.; Kealy, A.; Leach, J. Spatially enabling 'Place' information. Spatially Enabling Society: Research, Emerging Trends, and Critical Assessment; Rajabifard, A., Ed. GSDI Association, 2010.
18. Bennett, B.; Agarwal, P. Semantic Categories Underlying the Meaning of 'Place'. Conference on Spatial Information Theory; Winter, S.; Duckham, M.; Kulik, L.; Kuipers, B., Eds. Springer, 2007, Vol. 4736, *Lecture Notes in Computer Science*, pp. 78–95.
19. Winter, S.; Freksa, C. Approaching the notion of place by contrast. *Journal of Spatial Information Science* **2012**, *5*, 31–50.
20. Tuan, Y.F. *Space and Place : the perspective of experience*; University of Minnesota Press: Minneapolis, 1977.
21. Cresswell, T. *Place: An Introduction*; John Wiley & Sons, 2014.
22. Agarwal, P. Operationalising Sense of Place'as a Cognitive Operator for Semantics in Place-Based Ontologies. Conference on Spatial Information Theory. Springer, 2005, Vol. 3693, pp. 96–114.
23. Curry, M.R. *The work in the world: geographical practice and the written word*; U of Minnesota Press, 1996.
24. Agnew, J. Space and place. *Handbook of geographical knowledge* **2011**, *2011*, 316–331.
25. Vasardani, M.; Winter, S.; Richter, K.F. Locating place names from place descriptions. *International Journal of Geographical Information Science* **2013**, *27*, 2509–2532.
26. Adams, B.; McKenzie, G. Inferring thematic places from spatially referenced natural language descriptions. In *Crowdsourcing geographic knowledge*; Springer, 2013; pp. 201–221.

27. Hill, L.L. Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints. Research and Advanced Technology for Digital Libraries; Borbinha, J.L.; Baker, T., Eds. Springer, 2000, Vol. 1923, *Lecture Notes in Computer Science*, pp. 280–290.
28. Goodchild, M.F.; Hill, L.L. Introduction to digital gazetteer research. *International Journal of Geographical Information Science* **2008**, *22*, 1039–1044.
29. Silva, M.J.; Martins, B.; Chaves, M.; Afonso, A.P.; Cardoso, N. Adding geographic scopes to web resources. *Computers, Environment and Urban Systems* **2006**, *30*, 378–399.
30. Purves, R.S.; Clough, P.; Jones, C.B.; Arampatzis, A.; Bucher, B.; Finch, D.; Fu, G.; Joho, H.; Syed, A.K.; Vaid, S.; others. The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the Internet. *International Journal of Geographical Information Science* **2007**, *21*, 717–745.
31. Jones, C.B.; Purves, R.S. Geographical information retrieval. *International Journal of Geographical Information Science* **2008**, *22*, 219–228.
32. Montello, D.R.; Goodchild, M.F.; Gottsegen, J.; Fohl, P. Where's downtown?: Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition & Computation* **2003**, *3*, 185–204.
33. Jones, C.B.; Purves, R.S.; Clough, P.D.; Joho, H. Modelling vague places with knowledge from the Web. *International Journal of Geographical Information Science* **2008**, *22*, 1045–1065.
34. Grothe, C.; Schaab, J. Automated footprint generation from geotags with kernel density estimation and support vector machines. *Spatial Cognition & Computation* **2009**, *9*, 195–211.
35. Hollenstein, L.; Purves, R. Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science* **2010**, *2010*, 21–48.
36. Gao, S.; Janowicz, K.; Montello, D.R.; Hu, Y.; Yang, J.A.; McKenzie, G.; Ju, Y.; Gong, L.; Adams, B.; Yan, B. A data-synthesis-driven method for detecting and extracting vague cognitive regions. *International Journal of Geographical Information Science* **2017**, *31*, 1245–1271.
37. Speriosu, M.; Brown, T.; Moon, T.; Baldrige, J.; Erk, K. Connecting language and geography with region-topic models. Proc. Workshop on Computational Models of Spatial Language Interpretation (COSLI), 2010.
38. Adams, B.; Janowicz, K. Thematic signatures for cleansing and enriching place-related linked data. *International Journal of Geographical Information Science* **2015**, *29*, 556–579.
39. Twigger-Ross, C.L.; Uzzell, D.L. Place and identity processes. *Journal of environmental psychology* **1996**, *16*, 205–220.
40. Vasardani, M.; Stirling, L.F.; Winter, S. The preposition at from a spatial language, cognition, and information systems perspective. *Semantics and Pragmatics* **2017**, *10*.
41. Jordan, T.; Raubal, M.; Gartrell, B.; Egenhofer, M. An affordance-based model of place in GIS. 8th Int. Symposium on Spatial Data Handling, 1998, Vol. 98, pp. 98–109.
42. Kuhn, W. Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science* **2001**, *15*, 613–631.
43. Scheider, S.; Kuhn, W. Affordance-based categorization of road network data using a grounded theory of channel networks. *International Journal of Geographical Information Science* **2010**, *24*, 1249–1267.
44. Janowicz, K.; Keßler, C. The role of ontology in improving gazetteer interaction. *International Journal of Geographical Information Science* **2008**, *22*, 1129–1157.
45. Jarvella, R.J.; Klein, W. *Speech, place, and action: Studies of deixis and related topics*; John Wiley & Sons, 1982.
46. Couclelis, H.; Golledge, R.G.; Gale, N.; Tobler, W. Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology* **1987**, *7*, 99–122.
47. Richter, D.; Winter, S.; Richter, K.F.; Stirling, L. Granularity of locations referred to by place descriptions. *Computers, Environment and Urban Systems* **2013**, *41*, 88–99.
48. Bateman, J.A.; Hois, J.; Ross, R.; Tenbrink, T. A linguistic ontology of space for natural language processing. *Artificial Intelligence* **2010**, *174*, 1027–1071.
49. Mani, I.; Doran, C.; Harris, D.; Hitzeman, J.; Quimby, R.; Richer, J.; Wellner, B.; Mardis, S.; Clancy, S. SpatialML: annotation scheme, resources, and evaluation. *Language Resources and Evaluation* **2010**, *44*, 263–280.

50. Chen, H.; Vasardani, M.; Winter, S. Georeferencing places from collective human descriptions using place graphs. *Journal of Spatial Information Science* **2018**. Submitted and Available at: <http://www.josis.org/index.php/josis/article/viewFile/417/214>.
51. Kim, J.; Vasardani, M.; Winter, S. From descriptions to depictions: A dynamic sketch map drawing strategy. *Spatial Cognition & Computation* **2016**, *16*, 29–53.
52. Wolter, D.; Yousaf, M. Context and Vagueness in Automated Interpretation of Place Description: A Computational Model. *International Conference on Spatial Information Theory*. Springer, 2017, pp. 137–142.
53. Scheider, S.; Janowicz, K. Place reference systems. *Applied Ontology* **2014**, *9*, 97–127.
54. Alazzawi, A.N.; Abdelmoty, A.I.; Jones, C.B. What can I do there? Towards the automatic discovery of place-related services and activities. *International Journal of Geographical Information Science* **2012**, *26*, 345–364.
55. Janowicz, K. Sim-DL: Towards a Semantic Similarity Measurement Theory for the Description Logic ALCNR\mathcal{ALCNR} in Geographic Information Retrieval. On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops. Springer, 2006, pp. 1681–1692.
56. Hobel, H.; Fogliaroni, P. Extracting semantics of places from user generated content. *Proceedings of the 19th AGILE international conference on geographic information science* Google Scholar, 2016.
57. Levinson, S.C. Frames of reference and Molyneux's question: Crosslinguistic evidence. *Language and space* **1996**, *109*, 169.
58. Tenbrink, T. Reference frames of space and time in language. *Journal of Pragmatics* **2011**, *43*, 704–722.
59. Wallgrün, J.O.; Frommberger, L.; Wolter, D.; Dylla, F.; Freksa, C. Qualitative spatial representation and reasoning in the SparQ-toolbox. *Spatial Cognition*. Springer, 2006, Vol. 4387, pp. 39–58.
60. Cohn, A.G.; Renz, J. Qualitative spatial representation and reasoning. *Foundations of Artificial Intelligence* **2008**, *3*, 551–596.
61. Lynch, K. *The image of the city*; Vol. 11, MIT press, 1960.
62. Tomko, M.; Winter, S. Describing the Functional Spatial Structure of Urban Environments. *Computers, Environment and Urban Systems* **2013**, *41*, 177–187.
63. Moncla, L.; Gaio, M.; Nogueras-Iso, J.; Mustière, S. Reconstruction of itineraries from annotated text with an informed spanning tree algorithm. *International Journal of Geographical Information Science* **2016**, *30*, 1137–1160.
64. Yao, X.; Thill, J.C. How Far Is Too Far?—A Statistical Approach to Context-contingent Proximity Modeling. *Transactions in GIS* **2005**, *9*, 157–178.
65. Richter, D.; Winter, S.; Richter, K.F.; Stirling, L. How people describe their place: Identifying predominant types of place descriptions. *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*, 2012, pp. 30–37.
66. Kuhn, W. Core concepts of spatial information for transdisciplinary research. *International Journal of Geographical Information Science* **2012**, *26*, 2267–2276.