

ONTOLOGY-BASED SEMANTIC MODELLING OF PLACE KNOWLEDGE

by

DOORI OH

(Under the Direction of Xiaobai Angela Yao)

ABSTRACT

Place is an abstract concept that is difficult to describe or define comprehensively and objectively. There is a wide gap between how people describe places in daily life and places represented in databases in terms of data construction. Traditionally, places are recorded by authoritative agencies using a simple structure with the dataset of place name, its type, and footprints. For the last few years, as spatial data gets more attention in the field of GIS (Geographic Information Systems), there is a need to represent knowledge about places more specifically in terms of data semantics. This dissertation aims to propose an organizational architecture of place concepts that combines elements from multiple data sources using a computational ontological approach.

It is important to describe how traditional and non-traditional data sources can represent place information in an ontology. By comparing different concepts and the relations used in various sources, this study intends to outline the consistent components that can be used in a place ontology. To achieve this, traditional gazetteers such as the Geographic Names Information System (GNIS), GeoNames, the Getty Thesaurus of Geographic Names (TGN), and non-traditional data sources such as Google Places and Twitter data, were examined to identify the most often utilized concepts and relations

when demarcating place information. As a pilot study, Twitter data was used to explore people's activities for a better understanding of how people utilize places, and to assess the semantics among place categories.

Finally, place information extracted from various data sources is represented in ontological syntax. GeoSPARQL standard developed by the Open Geospatial Consortium (OGC) was reused in the ontology to describe geometries and the properties of spatial relations among places. Findings in this study, including a summary of comparisons with existing gazetteer ontologies, demonstrate the potential for using expert and non-expert (e.g., social media users) knowledge to enrich place knowledge. Thus, this dissertation contributes to discourses on digital gazetteers and geographic information retrieval (GIR). The framework of this study can be generally applied to other web-based platial data to increase information accessibility about places with a wide range of semantics.

INDEX WORDS: geospatial semantics, geographic information retrieval, ontology, gazetteers, platial data, GIS

ONTOLOGY-BASED SEMANTIC MODELLING OF PLACE KNOWLEDGE

by

DOORI OH

BS, Kyung Hee University, South Korea, 2011

MS, Kyung Hee University, South Korea, 2013

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

© 2021

Doori Oh

All Rights Reserved

ONTOLOGY-BASED SEMANTIC MODELLING OF PLACE KNOWLEDGE

by

DOORI OH

Major Professor:	Xiaobai Angela Yao
Committee:	Lan Mu
	Marguerite Madden
	Ismailcem Budak Arpinar

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2021

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my advisor, Dr. Xiaobai Angela Yao, and the members of my committee, Dr. Lan Mu, Dr. Marguerite Madden, and Dr. Ismailcem Budak Arpinar, for devoting their time and guidance over the years. I would also like to thank Dr. Ugur Kursuncu for his scientific discussions and recommendations, which were always helpful. I am very thankful to my colleagues in Dr. Yao's group for their perspectives and feedback on my research. Being able to work with intelligent, creative, and thoughtful colleagues is truly what has made the graduate student chapter of my life fruitful and enjoyable. In addition, I would like to thank Dr. Jerry Shannon for his considerable encouragement and constant help with the entire Ph.D. process. Through his guidance, I have expanded my knowledge of geography and learned his attitude of understanding matters in multiple ways. I must also acknowledge all the friends that I made along the way. I would like to thank them for their priceless friendship and for making the entire process flow beautifully. Finally, I would like to thank my family for their continuous support and love. I would not have been able to complete my doctoral program without them.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER	
1 INTRODUCTION	1
Motivation.....	2
Research Questions and Objectives	6
Significance of the Study	8
2 THEORETICAL FRAMEWORK AND THE CONSIDERTATIONS FOR ONTOLOGY CONSTRUCTION	9
Introduction.....	9
Literature Review.....	9
Terminological Clarifications	13
Conceptual Place-Ontology Development.....	15
The Final Considerations for Ontology Construction.....	17
3 A CONCEPTUAL ONTOLOGY CONTSTRUCTION FOR TRADITIONAL AND NON-TRADITIONAL GAZETTEERS	19
Introduction.....	19
Data Sources	21

Concepts or Classes	25
Relationships.....	28
Results.....	29
Comparison and Discussion.....	32
Conclusion	36
4 ENRICHED KNOWLEDGE ABOUT NAMED PLACES USING PLACE DESCRIPTIONS EXTRACTED FROM TWITTER DATA	39
Introduction.....	39
Related Work	44
Proposed Approach.....	47
Case Study	50
Results and Discussion	59
Conclusion	66
5 CONCLUSION.....	69
BIBLIOGRAPHY.....	72

APPENDICES

A The developed ontology in OWL	81
B A total list of three factors (place affordances, events, key-descriptors) identified from Twitter data	81

LIST OF TABLES

	Page
Table 2.1: Ontology requirements specification.....	16
Table 3.1: The number of filtered tweets using the names of schools and colleges.....	24
Table 3.2: Latent topics related to schools and colleges.....	24
Table 3.3: Some relations described in the place ontology	29
Table 3.4: A comparison of existing gazetteer ontologies.....	35
Table 4.1: Summary of statistical features of place categories in Athens, GA, USA	51
Table 4.2: Examples of word sets forming topics for the place categories ADM2 and Stadium	54
Table 4.3: Overall similarity scores for distinct place categories with the individual similarity scores for three factors.....	56
Table 4.4: The number of instances for selected place categories in Athens, GA, USA ..	62

LIST OF FIGURES

	Page
Figure 2.1: Resource description framework graph.....	14
Figure 3.1: Conceptual model of the place ontology	21
Figure 3.2: The general architecture of the proposed approach for Twitter data analysis.	23
Figure 3.3: A fragment of the possible knowledge representation using the place ontology	31
Figure 4.1: The number of tweets collected for each place type in the target study area (Athens, GA, USA, on the left-hand side)	53
Figure 4.2: Cosine similarities between place categories by individual factors: place affordances, key-descriptors (top, left to right), and events (bottom)	55
Figure 4.3: Multiplicity of place types vs activity types.....	59
Figure 4.4: Multiplicity of place types vs event types	60
Figure 4.5: Multidimensional scaling (MDS) map for POI, Populated Place, and School	64
Figure 4.6.: Multidimensional scaling (MDS) map for Church and Neighborhood.....	65
Figure 4.7.: Place types matching between GNIS and GeoNames (left), and GeoNames and Google Places (right) based on the thematic characteristics	66

CHAPTER 1

INTRODUCTION

Place-name databases are designed to provide georeferenced pieces of information about places and the semantic connections between them. They exist either in the form of traditional gazetteers or non-traditional data repositories. Traditional gazetteers have been developed to record named places and their associated information, mainly related to place types and spatial footprints, using authoritative sources (Hill, 2000; Goodchild & Hill, 2008; Keßler et al., 2009). Using these elements of named places, traditional gazetteers have at least two data translation functions. One is from place name to location (e.g., to answer the question, where is the University of Georgia?), and the other is from place name to place type translation (e.g., to answer the question, what is Jittery Joe's Coffee?). More recently, non-traditional repositories of named places have been developed as collaborative databases that provide unofficial or colloquial descriptions of places (e.g., Google Places or OpenStreetMap [OSM]). Non-traditional data sources provide additional information about named places, such as those with multiple place names or multiple functionalities (thus multiple place types) for the same place location.

The gazetteer databases created by government agencies, commercial companies, and groups of volunteers are used to find places along with their attributes in web-based applications such as mapping, georeferencing, wayfinding, and Geographic Information Retrieval (GIR) services (Goodchild, 2004; Hill, 2009). Though they are widely used in

many scientific and commercial fields, there is little sharing of ontologies and semantic connections between different data sources (Popescu et al., 2008; Machado et al., 2011). Therefore, following the previous discourses on digital gazetteer ontologies (Regalia et al., 2018; Codescu et al., 2011), this dissertation proposes an organizational architecture of place references for multiple data sources using an ontological approach.

The first chapter introduces the main concepts. The second chapter focuses on the theoretical background of this dissertation and the specification of ontology requirements. The third chapter develops a conceptual framework of ontology for traditional and non-traditional data sources of named places, followed by the fourth chapter, which explores how non-traditional data sources can be applied to help people better understand the dynamic aspects of places. The fifth chapter details the conclusions of this research.

Motivation

Digital gazetteers play an essential role in GIR. Traditionally, gazetteers are built and maintained by agencies. Traditional gazetteers often include limited information about place names, describing them with predetermined types of attribute data, such as place types and spatial footprints. With these attributes, traditional gazetteers can be used to query information about places, often in conjunction with a map. Several authoritative gazetteers are widely known, including the Geographic Names Information System (GNIS) released by the United States Geological Survey (USGS), the Alexandria Digital Library (ADL) Gazetteer created by the University of California Santa Barbara (Hill, Frew, & Zheng, 1999; Goodchild, 2004), and the Getty Thesaurus of Geographical Names (TGN) developed by the Getty Research Institute.

Nowadays, some gazetteers are constructed using a data-driven approach. For example, GeoNames is a global geographical database that maintains data from official sources and data contributed from volunteers. Thus, GeoNames is a large dataset, which contains a worldwide list of named places. Other examples include a collaborative data repository, such as OSM and Google Places. OSM provides a large dataset of tags created by users to represent attributes of places. Due to its openness and flexibility, OSM has huge potential to enrich place data. Google Places is designed to provide local business information and incorporates user-contributed information, which includes unofficial and colloquial descriptions of places.

Several limitations are found in existing gazetteers. First, traditional gazetteers often include specific yet limited information about place names, and the scope may also be restricted to a particular time frame (Southall et al., 2011). Place names are described with predetermined types of attribute data, such as place types and spatial footprints. However, named places reveal more than predefined attributes of places, depending on the people who use them and how they choose to use them (Golledge, 2002; Janowicz et al., 2009). Therefore, it is necessary to expand the contents in traditional gazetteers to include up-to-date and practical information about places by adding new concepts and relationships (e.g., the occasional use of a place).

Secondly, to further problematize the use of different attributes to represent a place, how places are conceptualized can differ according to the viewpoint through which one characterizes them (Janowicz & Keßler, 2008; Keßler et al., 2009). For example, a place can be classified as a building in one gazetteer and as a museum in another gazetteer. Place types such as *point of interest* (POI) or *building* are commonly used with

other place types in gazetteers. Consequently, users may find different types of information about a place name from various data sources.

Finally, the functional aspects of places are often neglected in traditional gazetteers (e.g., a bar can be used for a concert, a chapel can be used for a wedding). The meanings of place names evolve over time, and so do the meanings associated with the places. The place name, and therefore the place it refers to, is not only provided from local knowledge about that place at a time but also evolves dynamically with people and their experiences, which are often manifested on social media platforms. For example, Twitter users may post a wide range of content about places associated with social events in a specific time period. Therefore, it is essential to examine place descriptions generated by the public to enrich and update the existing place databases.

The current study is motivated by believing that places' static and dynamic properties are all important in the comprehensive knowledge about named places. In response to the limitations of current place-name databases, this study calls for an ontology-based approach that integrates core concepts of gazetteer information from various data sources, as well as additional references to place names that can be extracted from social media data. Therefore, the proposed place ontology reveals more comprehensive knowledge about places than the gazetteer information alone, so this framework may help populate place names with enriched information from a wide variety of data sources.

The following two arguments elaborate the rationale for an ontology-based approach to representing schema for gazetteer information. First, gazetteers categorize information about place names according to a classification scheme; this enables one to

answer questions such as ‘What school names are in Athens, GA?’ within GIR systems. However, there is a lack of explicit definitions for terms and semantic relationships shared among different data sources (Popescu et al., 2008; Machado et al., 2011). Users have to interpret the terms used in different sources based on metadata. Moreover, some concepts and relationships may be missing in a gazetteer. Therefore, an ontology that includes definitions of a comprehensive set of concepts and relationships is desirable for successful knowledge representation and reasoning about place names and associated place information.

Secondly, the concept of ontology in this study is based on a computer science perspective, in which ontology is considered ‘logical theory accounting for the intended meaning of a formal vocabulary’ (Guarino, 1998). This definition refers specifically to a process that allows a computer system to connect semantically related knowledge (Gruber, 1993; Studer et al., 1998; Sowa 2000). In terms of gazetteer data management, ontologies have been developed by the USGS¹, the UK Ordnance Survey², GeoNames³, Getty Research Institute⁴, and OSM⁵, to name a few examples. Most gazetteer ontologies aim to improve accessibility across their databases. In contrast, the present research was conducted to develop an ontology in response to the need to utilize different data sources that provide comprehensive information about named places.

¹ USGS, <https://www.usgs.gov>

² Ordnance Survey, <https://ordnancesurvey.co.uk>

³ GeoNames, <https://www.geonames.org>

⁴ Getty Research Institute, <https://www.getty.edu/research>

⁵ OpenStreetMap, <https://www.openstreetmap.org>

Research Questions and Objectives

The overarching question that this dissertation attempts to answer is **how to enrich place information for better data management and analysis?** To answer the overarching question, three research objectives were constructed.

1. **To integrate the descriptive concepts and subconcepts of named places in the existing traditional gazetteers and non-traditional place repositories.**

Describing richly detailed place information presents many challenges. In particular, gazetteers are built for different purposes, with various scopes of data. Through this study, the concepts and taxonomies commonly used in multiple data sources of named places were examined. Thus, this research builds on the previous system of geographic features and provides the potential for enriched knowledge about places. The following questions are addressed in order to achieve this objective.

- a. What types of concepts and taxonomy are commonly represented in traditional and non-traditional gazetteers?
- b. How can extracted concepts that provide place descriptions be integrated into a consistent knowledge base?

2. **To investigate people's activities related to named places that can be utilized to enrich a place categorization system.** Though existing gazetteers might be sufficient to specify the relationships between place names and locations, places can be defined in many other ways. It is appropriate that places are interpreted as spaces where people interact with locations. In other words, a place has phenomenological meanings. The current research emphasizes the functional

aspects of named places using social media data. The following questions of capturing the recognized place information by social media users were answered to develop a schema describing a set of concepts, also known as classes, of places in study areas.

- a. What type of concepts can be captured from social media data when obtaining information about people's activities related to named places?
- b. How can concepts captured from social media data be used to better understand gazetteer components?

3. **To construct a place ontology for traditional and non-traditional data**

sources of named places. A place may be multifaceted both in its function and in human experiences of it. Although places can be interpreted through individual experiences, there are always types of idea that most people can relate to. The proposed place ontology contains official data about named places and knowledge about places extracted from social media data. The purpose of this ontology is to find places matching questions such as 'the locations of places to go running that have trails and gardens in Athens, Georgia.' To achieve this objective, the following questions were asked.

- a. What are the specific requirements for the ontology?
- b. How can elements of named places extracted from multiple data sources be integrated into the proposed place ontology?
- c. How can spatial relations between places be represented in a place ontology?
- d. What are the differences between the proposed place ontology and the existing gazetteer ontologies?

Significance of the Study

The intellectual merit of this dissertation lies in its unique contribution to the discourses on digital gazetteers and GIR. Until recently, gazetteer ontology research emphasized formal traditional place information and largely ignored the context of human experience, including actual usage-pattern-based place information, as a domain. In order to construct a place ontology to retrieve comprehensive place information, this research proposes using social media data to extract the dynamic uses of places together with the traditional data sources of gazetteers. Therefore, ontological designs in this study include aligning the concepts of named places and relationships among them, focusing on their environmental characteristics and functional aspects. Furthermore, this study compares the proposed place ontology to the existing gazetteer ontologies to verify its improvement. Thus, the place ontology aims to engage with research questions in GIS, such as how best to describe information about named places using ontologies to enable sharing and reuse of knowledge and data.

In addition, the proposed project will contribute to the fields of urban studies and GIS. This study identifies a broader range of information about named places by incorporating both traditional data sources and user perspectives. Thus, the results of this dissertation will provide urban planners with a greater understanding of place information as a framework for their decision-making practices. Moreover, this project proposes necessary commitments on place information by expanding discourses about ontology-based knowledge construction. Thus, an ontological understanding of domain knowledge will help utilize relevant geographic information for knowledge modeling and decision support.

CHAPTER 2

THEORETICAL FRAMEWORK AND THE CONSIDERATIONS FOR ONTOLOGY CONSTRUCTION

Introduction

This chapter presents and discusses the previous ontology-based works on gazetteers, terminological clarifications, and the requirement specifications for ontology design and development. These ontology requirement specifications and the rules for matching similar concepts used in different place-name databases are then used for conceptual ontology construction discussed in Chapter 3.

Literature Review

While the notion of place in the field of geography has been defined in various ways, most attempts at defining the concept have taken a space-centric perspective (Janowicz, 2009; Cresswell, 2014). Places are identified via given names and are categorized according to commonly agreed-upon characteristics (Hill, 2009). Places can also be referred to by descriptions shared within a community (e.g., North Georgia is often used to describe the region north of the Atlanta metropolitan area in the USA). Even if the geographic location of a place tends to be relatively stable, its functional definition is subject to change over time. Therefore, the meaning of places encompasses a wide range of functions resulting from interactions between individuals and a specific location (Janowicz, 2009; Goodchild, 2011; Zhong et al., 2014).

Ontologies has grown in AI and information systems pragmatically for domain knowledge representation and knowledge-based problem solving (O'Leary, 1998; Chandrasekaran et al., 1999). This is motivated by the need for commitments to modeling domain knowledge. Therefore, ontologies enable coherent reasoning processes in information systems. From AI perspectives, ontologies are used to denote knowledge engineering for ontology-driven information systems in GIS (Mark et al., 2004; Smith & Mark, 2001, 2003). In particular, geographic information scientists (GIScientists) develop and apply geo-ontologies to formalize geographic concepts in GIS terms (Fonseca et al., 2000; Frank, 2001). As a critical source of geographical information, gazetteers play a crucial role in GIS. In previous studies, infrastructures for ontology-based gazetteers have been developed to improve the retrieval of information about places (Ballatore, 2016). For example, the USGS and the Ordnance Survey in the UK have developed their gazetteer ontologies for the national maps in their respective countries (Regalia et al., 2018).

Furthermore, some ontology-based works aim to recognize and extract knowledge about places from web-based services. OnLocus was developed to describe spatial and semantic relationships between locations using geographic knowledge extracted from web pages (Borges et al., 2007; Borges et al., 2011). OnLocus uses the extracted positioning expressions, such as postal codes or telephone area codes, to refer to places. Toponym ontology (TO) has been developed to identify equivalent toponyms using several web resources (Smart et al., 2010). The primary purpose of TO is to integrate distributed gazetteer sources. In many cases, gazetteer ontologies have been developed

individually as single systems, and the need for ontology alignment has been widely recognized for defining standardized terminologies (Cruz et al., 2004).

The process of ontology alignment involves a wide variety of subjects and techniques. Sunna and Cruz (2007) introduced a centralized architecture and a peer-to-peer architecture for this alignment. The former aligns the associated ontologies to the global ontology, whereas the latter establishes links between two ontologies. In both cases, the source and target ontologies are translated into a third ontology. The matching approaches can be classified into element-level, structure-level, and hybrid approaches (Shvaiko & Euzenat, 2005; Sunna & Cruz, 2007). The element-level ontology alignment applies to individual elements in an ontology, such as concepts and relationships. In contrast, the structure-level techniques examine combinations of elements and focus on the structure of ontologies. Hybrid approaches combine different methods to align ontologies.

There have also been studies about instance-based ontology alignment, which exploit the similarity between common instances to discover the semantic relationships between concepts in two ontologies (Brauner et al., 2007). Other approaches require the use of external sources, such as dictionaries and thesauri, to interpret the meaning of concepts and relationships in an ontology (Lambrix & Tan, 2006). In light of findings from previous studies, the proposed conceptual framework in this dissertation has utilized the hybrid approaches, matching multiple data sources based on the synonyms and definitions of concepts and relationships (i.e., element-level methods) and the similarity of concepts in their environment, such as the parents and the children of a concept (i.e., structure-level approaches).

Another important aspect is that ontology alignment is often an essential step for data integration. In terms of data integration, Manguinhas et al. (2008) presented a geo-temporal gazetteer service integrating data from multiple sources (e.g., files or records from the national authorities, GeoNames, the GeoNetPT web ontology language [OWL] ontology, or other web resources). Moura and Davis (2014) proposed a gazetteer enriched with semantic relationships and connections between non-geographical entities. Their work combined two linked data sources of gazetteer data, namely GeoNames and DBPedia, to populate an integrated and semantically enriched gazetteer. Some works have focused specifically on the alignment of feature types, which is often regarded as an obstacle to integrating gazetteers (Brauner et al., 2007; Janowicz & Keßler, 2008). This dissertation proposes a conceptual framework that embraces traditional and non-traditional gazetteers, along with place descriptions extracted from social media data. Thus, this approach can help to align information about place names from a wide range of sources.

Due to the explosion of place information available online, many studies have used external web resources to expand gazetteers (Vasardani et al., 2013). The most popular web resources used for gazetteer expansion are Wikipedia (Kazama & Torisawa, 2007; Toral & Muñoz, 2006; Zhang & Iria, 2009), geo-coded data (Gelernter et al., 2013), and social media data (Zhang et al., 2021). Unlike other studies, the dissertation research aims to build additional attributes of places that can retrieve place names in the place ontology, not to try to extract place names from web resources. Thus, this research contributes to enriching a gazetteer by seeking a way to link multiple data sources on the conceptual level.

Terminological Clarifications

For clarification, some important terms used in this dissertation are described below. These descriptions include three web ontology languages and their components. The World Wide Web Consortium (W3C)⁶ defines a layered architecture of a semantic web, which consists of languages and technologies (e.g., resource description framework [RDF], RDF schema [RDFS], and OWL) that specify the semantic properties of web resources. To represent integrated knowledge about named places from multiple data sources, this dissertation used the concept of ontologies and ontology languages to describe the proposed conceptual place ontology.

- RDF⁷ is a data model for describing resources on the web, agreed upon by the W3C. RDF is designed in a computer-processable format, which can be represented on a graph, and comprises a subject, predicate, and object (Figure 2.1).
- RDF Schema⁸ is a simple ontology language (compared with OWL) used to define classes, instances, and properties in an RDF graph. In other words, RDF represents data in the form of triple statements, and RDFS describes the semantics in RDF descriptions.

⁶ W3C stands for World Wide Web Consortium, which is the de facto standardization organization for worldwide-web-related technologies (<https://www.w3.org/>).

⁷ <https://www.w3.org/RDF/>

⁸ <https://www.w3.org/2001/sw/wiki/RDFS>

- OWL⁹ extends RDF and RDFS to provide enough constructs and restrictions for semantic information. OWL semantically extends RDFS, so most of the RDFS components are included in the corresponding OWL. In the current study, the proposed place ontology was encoded in Protégé 5.5 for OWL/RDFS ontologies in Chapter 3.
- ‘Individuals’ represent the ground-level components of an ontology (e.g., specific place names). Individuals are also called ‘instances,’ and they are placed into a class.
- ‘Classes’ and ‘concepts’ could be used interchangeably. A class represents all individuals or instances that share the same attributes within the shared knowledge domain.
- ‘Relationships’ are applied to classes or individuals to specify how one is related to another. They can also be called ‘properties’ in a specific environment (e.g., in OWL). For example, the class ‘place name’ has a relationship with the class ‘place type’.

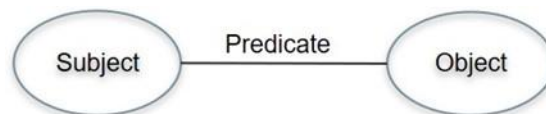


Figure 2.1: Resource description framework graph

⁹ <https://www.w3.org/OWL/>

Conceptual Place-Ontology Development

The requirement specification of the proposed place ontology is presented in Table 2.1. The primary purpose of building a place ontology is to enrich the semantics of concepts and the conceptual relationships forming knowledge about place names. This study focuses on the official information about place names proposed by authorities and the practical use of named places by people. For example, the University of Georgia has many subtype places within schools, such as ‘garden,’ ‘chapel,’ ‘museum,’ etc. This research aims to represent these informal descriptions about named places based on the frequency of mentions by people along with official data from the gazetteer databases. These concepts and the relations among these concepts are linked using ontology-based modeling.

The core terms and taxonomies across the gazetteer databases are interpreted to combine them into an ontology. This dissertation used the definitions of the concepts to investigate any similar uses. For example, the concept ‘place designation’ and its subconcepts are described in the GNIS metadata separately from other core concepts. They could be incorporated into more generic taxonomies of the concept ‘place type.’ In the place ontology, different levels of granularity (Fonseca et al., 2002) can be applied to hierarchical concepts and taxonomies. All elements were defined in semi-formal expressions for conceptual-level ontology and represented using formal terms in the computational environment.

The resulting place ontology was designed to provide a consistent knowledge base. It was specifically targeted for knowledge representation of places. Geospatial software developers can also use it for providing ontology-based gazetteer services. The

intended uses of the place ontology were to store and manage place names and information about places consistently. The proposed ontology contains extensive knowledge of named places. It can be used to retrieve place data using various search terms (e.g., matching criteria based on official and unofficial information about places).

Table 2.1: Ontology requirements specification

	Ontology requirements	Specification
1	Domain	Conceptual and practical knowledge of place
2	Purpose	Conceptualization and formal representation of knowledge about places for diverse data sources
3	Extent of Knowledge	Place attributes of places limited for the traditional gazetteer, non-traditional gazetteer, and the knowledge of places that can describe the usage of places
4	Level of Formality	Semi-formal expressions for conceptual ontology, Formal terms for logical ontology
5	Granularity	The same level of granularity in the data structures of multiple gazetteers
6	Intended End-Users	User 1. Researchers and domain experts seeking to understand the knowledge representation of places User 2. Geospatial software developers for providing ontology-based services
7	Intended Uses	Use 1. Store and manage place data Use 2. Propose a service in which the users can find general information about a place, its location, and other information Use 3. Provide statistics and recommendations of places using search terms
8	Source of Knowledge	The specified knowledge in the GNIS, GeoNames, TGN, Google Places, and human perceptions on places that can be found in Twitter posts

This dissertation used traditional gazetteers (i.e., GNIS, GeoNames, and TGN), a non-traditional place name repository (i.e., Google Places), and an external web resource (i.e., social media data) to extract terms and taxonomies of named places. The place ontology could be expanded in further studies using other data sources such as digital news or blogs.

The Final Considerations for Ontology Construction

It is essential to clarify the items in the domain to be modeled when building an ontology (Van Assem et al. 2004). This dissertation proposed several rules to support ontology construction with consistency. The purpose of the suggested rules was to ensure that the concepts and relationships were included in the specified scope and purpose of the ontology.

- Rule 1: A place ontology has a common superclass called ‘Place.’ All individuals of subclasses are in the class of ‘Place.’
- Rule 2: The concepts commonly used across the place databases and conceptual concepts explicitly associated with the domain knowledge are included into an ontology. Concepts created for a specific application are excluded. However, these concepts can be specified in a local ontology to provide additional information for a particular community of users.
- Rule 3: If the definition of a concept is not used consistently in different data sources, it will be extended until it is applicable for generic cases.

- Rule 4: If a concept is defined differently and these definitions contradict each other, new concepts are introduced to provide explicit definitions. In this case, new concepts will be named using one of the alternative terms.
- Rule 5: All concepts are defined with formal specifications. They are named in the singular form as a naming convention for concepts in a domain ontology.

In most cases, the existing place databases do not use taxonomic relationships. In addition, they do not distinguish between subsumption relationships (is-a, also called a generic hierarchical relationship). Therefore, it must be checked whether each class is related to its superclass in terms of inheritance. In the proposed ontology, a class was allowed to have any number of parent classes (multiple inheritance). For example, if a subclass was specified in two parent-classes, that subclass automatically showed the inherited properties from both parent classes. Any conflicting information dealing with multiple inheritances was handled through Rules 1–5.

CHAPTER 3

A CONCEPTUAL ONTOLOGY CONSTRUCTION FOR TRADITIONAL AND NON-TRADITIONAL GAZETTEERS

Introduction

In response to the limitations of the current place-name-based approach to place repositories, this pilot study extended the framework for place knowledge by developing an ontology that can accommodate place information from multiple sources, including (1) traditional gazetteers (i.e., the GNIS, GeoNames, and the Getty TGN); (2) non-traditional repositories (i.e., Google Places); and (3) place descriptions in social media data (i.e., Twitter data). This framework is open because the ontological framework was designed to take data from existing place repositories. More importantly, it was intended to be updated dynamically by retrieving the latest information about named places in the public domain. This ontological framework reused and integrated various relationships between places in the current data silos and from the emerging data sources based on archival research on existing place-name repositories. The simplified structure of the place ontology is depicted in Figure 3.1.

The proposed place ontology was built upon different types of relationships between places. Creating a place ontology requires exploring multiple gazetteer databases and existing gazetteer ontologies, ontological conceptual design, and computational implementation. Gazetteer metadata and relevant documents were reviewed to identify different perspectives in creating place-name databases, categorizing place-related

concepts from various data sources, and discovering semantic relationships among them. In addition, the study tested the feasibility of taking social media data by implementing a small pilot study. Twitter data were used to conduct a prototype study to build the relationships between pieces of information about place names. The pilot study collected and pre-processed tweets to make them suitable for the subsequent data mining of place information. Topic modeling, a form of text mining, was applied to analyze hidden topics appearing with place names in the pre-processed tweets. Extracted topics were labelled as the additional attributes of place names and stored to form place ontology. Some place-based events and relevant activities were discovered from Twitter data using a probabilistic topic model in a pilot study. The extracted concepts from Twitter data were then added to the proposed ontology using a bottom-up strategy. For example, many tweets were posted about sports games, such as football, mentioning school and college names. Therefore, the concept 'Sport Event' was added with a relationship to the concept 'School.' Thus, place information extracted from the tweets can be used to retrieve other information from gazetteer databases.

The main goal of this study was to implement a conceptual place ontology to facilitate the data integration of a variety of data sources, including traditional gazetteers, non-traditional place repositories, and place information that can be extracted from other resources through information retrieval (IR). The significance of this study is twofold. First, it identified and bridged minor differences in the shared semantics across data sources by suggesting a comprehensive ontology framework. Secondly, the proposed ontology extended the semantics of place descriptions from static to dynamic aspects of places. Updated information about dynamics aspects (i.e., practical uses) of places can be

retrieved from, for instance, social media data. Consequently, this study demonstrated how place information from multiple sources was incorporated into an ontology-based framework to maximize the chances of finding place names via a wide range of search criteria.

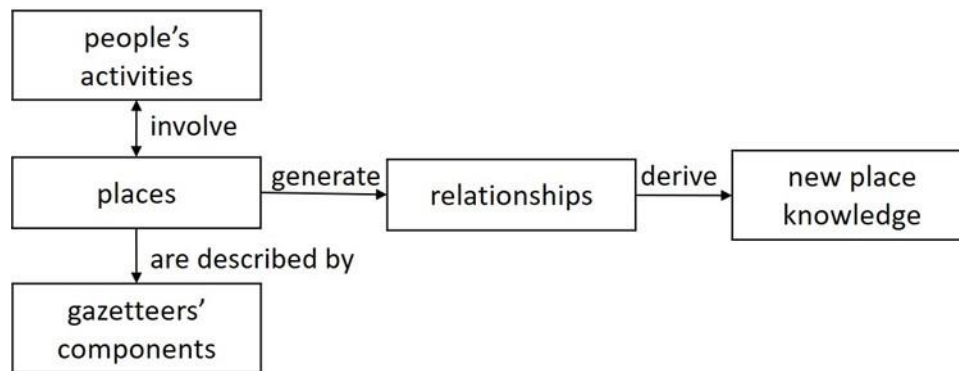


Figure 3.1: Conceptual model of the place ontology

Data Sources

The GNIS, released by the USGS, is the official repository of domestic geographic names in the USA and is used as a source of place names on the national map. GeoNames is a global geographical database that includes over 11 million place names worldwide, and it can be accessed via its web services. GeoNames users can contribute to the database by adding or editing the GeoNames data. The Getty TGN is another well-known gazetteer service and has a unique thesaural structure. As an example of an unofficial repository, Google Places provides place searching with a list of place types and detailed information about a specific location.

In addition to the above-discussed data sources, this study created ontology elements using Twitter data. For a pilot study, tweets were collected in Athens, GA, USA, using the Streaming Twitter application programming interface (API) between December 2017 and January 2019. The area was chosen because the researchers were familiar with the city and had solid contextual knowledge of the place information posted in tweets. The collected tweets were then filtered for a specific type of place. As part of the pilot study, this study focused on the education system, using a list of schools and colleges, and their building names. In the next step, tweets were analyzed to retrieve the types of activity or public events associated with the school names in the tweets. Such information helped to explain the functionality of the place type ‘School.’

Twitter posts were used to identify the possible concepts for describing places in a place ontology. We used topic modeling to discover the latest topics from tweets to create proper subconcepts of place knowledge. Topic models, such as the Latent Dirichlet Allocation (LDA) (Blei et al., 2003), are widely used to identify a latent topic structure in a corpus. The LDA is a generative model for describing the latent topics of documents and assumes that each document is represented as the word distributions of a topic (Wei & Croft, 2006). The original LDA was designed to identify hidden topics from documents such as news articles, academic papers, and blogs. Unlike the traditional documents, tweets are (1) short, with a 140-character limit; (2) include a mixture of short and informal words; and (3) include contextual elements, such as tags, user mentions, and URLs. Therefore, Twitter data were pooled for a specific type of place (i.e., school & college) to which the LDA model was applied.

The general architecture of the method is shown in Figure 3.2. The collected tweets were filtered using a list of schools, colleges, and their building names in the local area. Subsequently, the filtered tweets were grouped by season. Each dataset was pre-processed to remove the stop words (e.g., ‘a,’ ‘is,’ and ‘the’), URLs, non-alphanumeric characters, and Twitter mentions in Twitter posts. Only re-tweets by the same user were filtered out because re-tweets reflect the users’ interests. The pre-processed extracted tweets in each dataset are described in Table 3.1. LDA is conditioned on three parameters: document-topic distributions α , topic-word distributions β , and topic number K . In this study, they were set as $\alpha = 0.1$, $\beta = 0.05$, and $K = 50$. LDA was performed via 2000 iterations with the Gibbs sampling algorithm for each dataset. The topic coherence score was used to determine the quality of the learned topics. Table 3.2 lists the top topics in different time periods. These topics were labeled and stored as subclasses of ‘Function’, which were related to the class ‘School’ in a semantic database using RDF syntax in Section 4.

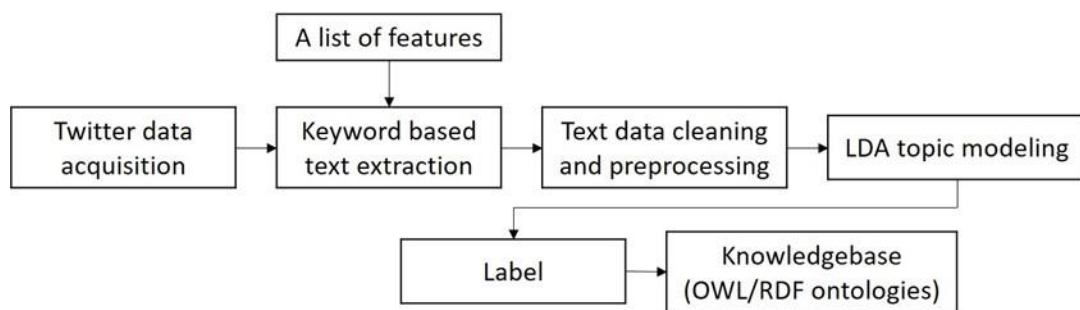


Figure 3.2: The general architecture of the proposed approach for Twitter data analysis

Table 3.1: The number of filtered tweets using the names of schools and colleges

Time period	Count	Sample tweet
Dec 2017 – Feb 2018	34,313	RT @KirbySmartUGA: #DawgNation we did it!! SEC Champs! Go Dawgs!
Mar 2018 – May 2018	26,079	Top 5% breakfast this morning. Another great group representing North Oconee High School!
Jun 2018 – Aug 2018	13,875	Here's a recap of this week's #NewFacultyTour. We welcome our new faculty members to the University of Georgia community and Franklin College...
Sep 2018 – Nov 2018	32,776	Weather over next 8 hours at UGA Butts-Mehre Building (https://t.co/tlZpow0eDN)
Dec 2018 – Jan 2019	30,438	RT @UGAfootballLive: What a rookie season for #UGA great Roquan Smith (129 tackles)

Table 3.2: Latent topics related to schools & colleges

Category	Topic	Associated words
Political event	Polling	Democrat, voters, media, etc.
Social event	College tour	Tour, students, campus visit, etc.
Sport event	College football game	SEC Championship, football, coach, etc.
Fun activity	Tailgating	Tailgating, dawgnation, godawgs, etc.
Residential activity	Dorming	Roommate, room, etc.
Service activity	Weather report	Forecasts, weather, butts-mehre building, etc.

We used several place name repositories and place descriptions from tweets to enrich the knowledge about places. Along with traditional approaches to representing places in the existing gazetteers, tweets are a source of updated information about places. The place ontology was used to classify place-related concepts and their relationships from the multiple data sources.

Concepts or Classes

To provide defined concepts for an ontology, this study examined the commonly adopted categories (i.e., place name, place type, and geographic location) to represent place names and information about places in the current place databases. In some cases, a concept was represented at different levels of detail depending on the system. For example, place names could have subconcepts such as ‘preferred name’, ‘alternative name’, or ‘historical name’ in one gazetteer while remaining as single concepts in another gazetteer. In this case, subconcepts were included in the ontology if their purposes and definitions were clarified in the current gazetteer databases.

In some cases, the taxonomy of a concept was inconsistently applied in different data sources. Of these concepts, place types were the most inconsistent concepts used in the gazetteers. For this research, the list and hierarchy of place types were identified using guidance from the United States National Geospatial-Intelligence Agency (NGA) and Google Places. The NGA provides geographical place types in thematic groups. NGA guidance or databases are widely used in the existing gazetteers, such as the GEOnet Names Server (GNS), GeoNames, and OSM. Google Places includes place types of urban environments, such as ‘gas station,’ ‘gym,’ ‘spa,’ ‘restaurant,’ or ‘shopping

mall.’ Therefore, the subtype relationships of place types represented in the proposed place ontology are more complex than a single gazetteer database.

Unlike the other concepts, ‘Function’ and their subconcepts were extracted from the tweets to present versatile or occasional uses of places in people’s daily lives. Functions are mainly associated with people’s activities and events at places. People’s activities such as learning, entertaining, eating, and relaxing are related to place affordances that are dynamic and reflect contexts based on surrounding environments. Events are non-continuous activities in places, which reflect occasional changes in terms of functionality and captures the functional characteristics from the perspectives of planners and practitioners. For that descriptive place information extracted from web resources can be managed to provide functional characteristics about places in the developed ontology.

In gazetteers, places are usually described as a single point, and most gazetteers do not represent their geometries (i.e., the objects’ shapes). To represent the geometries concept, GeoSPARQL standard concepts by the Open Geospatial Consortium (OGC) were reused. GeoSPARQL defines a set of terms for geospatial features with multiple geometries (e.g., multi geometry, curve, and triangle). The full descriptions of GeoSPARQL can be found on the OGC web page¹⁰. The proposed place ontology included six top-level concepts and their subordinate concepts based on place descriptions in various prominent data sources. The definition of each top-level concept is listed below. These concepts are formalized in Section 3.5.1.

¹⁰ GeoSPARQL - A Geographic Query Language for RDF Data, <https://www.ogc.org/standards/geosparql>

- Place: A geographic location designated for, or being used by, people. The class is a common superclass that defines several attributes for place names.
- Place name: A given name that indicates a place. The class has subclasses such as ‘Preferred Name,’ ‘Alternative Name,’ and ‘Historical Name.’
- Place type: An intended purpose of a place. The class has subclasses such as ‘Administrative Feature,’ ‘Area Feature,’ ‘Business Feature,’ ‘Hydrographic Feature,’ ‘Hypsographic Feature,’ ‘Populated Feature,’ ‘Spot Feature,’ ‘Transportation Feature,’ ‘Undersea Feature,’ and ‘Vegetation Feature.’
- Statistics: Descriptive statistics of survey data about a place (e.g., population and elevation). The class has subclasses such as ‘Demography’ and ‘Geography.’
- Function: How places are used and experienced in people’s daily life. It captures multi-functional aspects of a place to determine how people use places around them. The class has subclasses such as ‘Activity’ and ‘Event.’ In comparison, ‘Activity’ is associated with the action capabilities of individuals at places. While ‘Event’ is an occasional and large-scale occurrence at a place (e.g., local festival, marathon, and protest) that captures the functional characteristics from the perspectives of planners and practitioners.
- Spatial object: The geospatial representation of a place. The class is at the same level as the class ‘Place’ in the proposed ontology. It has subclasses

such as ‘Feature’ and ‘Geometry.’ These concepts are applied by reusing GeoSPARQL (The OGC standard).

Relationships

In the place ontology, classes and individuals can be related to one another using relationships. Therefore, relations between ontology elements can be used to query a knowledge base. Relations are represented as properties using intuitive names in the place ontology (Table 3.3). Relations between place names and relevant concepts (i.e., ‘hasPlaceName’, ‘hasPlaceType’, ‘hasFunction’, and ‘hasEvent’) support the retrieval of place information based on place names, types, functions, or events. For example, ‘hasEvent’ can be used to find place names related to an event. The relations for geometries and other spatial representations are defined in the GeoSPARQL standard. In gazetteers, a place is partly described using administrative regions where it is located (e.g., countries, states, and counties), and most gazetteers do not represent spatial relations between two places. Identifying spatial relations between two places using the current gazetteers requires additional knowledge and techniques, such as geospatial analytics and map references. To represent topological properties, including geometries in gazetteers (i.e., place locations usually described by a coordinates system), spatial relationships such as topological relationships are described to identify place information based on their spatial relationships with other places. For instance, ‘hasGeometry’ and ‘contain’ relations can be used to find places that spatially contain (or are contained by)

other places. As an example, the Coastal Grill restaurant is in Five Points (a landmark area) in Athens. These relations are formalized in Section 3.5.2.

Table 3.3: Some relations described in the place ontology

Core Relationship	Type	Description
hasPlaceName	Place × PlaceName	The various names given to a place
hasPlaceType	Place × PlaceType	The various types given to a place
hasFunction	Place × Function	Multi-functional aspects of a place
hasEvent	Place × Event	Events associated with a place
hasGeometry	Feature × Geometry	Geometries described in GeoSPARQL standard
Other representations of spatial relations (e.g., contains, crosses, disjoint, intersects, overlap, touches, within, etc.)	SpatialObject × SpatialObject	Spatial relations described in GeoSPARQL standard

Results

Classes

Following the proposed conceptual framework, a place ontology was implemented in Protégé 5.5, which is an open-source ontology editor developed by the Stanford University School of Medicine. The schema of place knowledge was

represented using OWL. The proposed place ontology provided semantic information about the place elements extracted from gazetteers and Twitter data, and geometries concepts from GeoSPARQL helped represent spatial relations among places. The proposed place ontology contained six top-level concepts and their subordinate concepts, with the concepts used in GeoSPARQL. The classes were arranged hierarchically. A total of 798 classes were created at different levels of granularity. The encoded ontology file can be found on the GitHub repository¹¹ and can be opened using the ontology editor Protégé.

Place information can be represented at different levels of granularity depending on how it was addressed in existing place-name databases. One class of the ontology may describe more details by specifying the attested subclasses, while another class maintains a simple structure without a subclass. For example, the class ‘Place Name’ included the three subclasses in the place ontology, whereas the class ‘Place Type’ had 10 subclasses, which then branched out into 686 subclasses. These classes were populated with individual places or relevant data (e.g., numerical data for statistics). In most cases, a place instance was a member of several classes. In ontologies, classes and subclasses have hierarchical relations (is-a) between them. Additionally, other class axioms specify necessary class information that could help to represent place knowledge efficiently. If two classes are semantically equivalent, they might have the same instances. In the place ontology, some equivalent classes existed when the terms could be used interchangeably

¹¹ The ontology file openly available in a public repository,

https://github.com/doh30/Geo_Ontologies/blob/main/gaz_onto_v1.0.owl

(e.g., ‘RailwayStation’ and ‘TrainStation’). The equivalent classes were encoded in the place ontology to combine place knowledge from multiple sources.

Relations

The term property used in the implementation stage referred to a relation between a set of individuals or between an individual and data value. The developed ontology included properties built from the existing gazetteers, Twitter data analysis, and GeoSPARQL standard. Properties were used to represent the relations between semantic information about places in the ontology. For example, a place was linked to a social event with the property ‘hasSocialEvent’ where ‘hasSocialEvent’ was the property that described the relationship between the place and the way that place was used. Figure 3.3 shows a fragment of the possible knowledge representation using the developed place ontology.

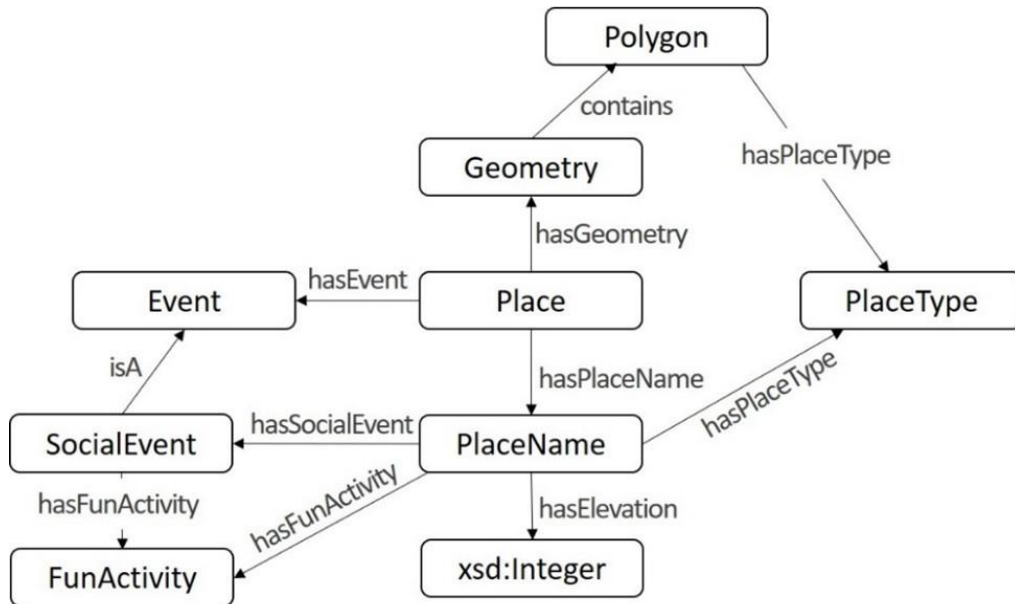


Figure 3.3: A fragment of the possible knowledge representation using the place ontology

One of the strengths of ontology modeling lies in knowledge reasoning through given properties and constraints. To make a knowledge representation more effective, the developed ontology defined the domain (i.e., the class to which a property was applied) and range (i.e., the class that a property had as its value) of properties, which are both features of ontology languages such as OWL. The properties ‘hasPlaceName,’ ‘hasPlaceType,’ and ‘hasFunction’ had the class ‘Place’ as their domain and the respective classes as their range (the class ‘PlaceName’, ‘PlaceType’, and ‘Function’). Spatial relations by the GeoSPARQL standard had the class ‘SpatialObject’ for both their domain and range. The property ‘hasGeometry’ had the class ‘Feature’ as its domain and ‘Geometry’ as its range.

Unlike most properties that link individuals to classes, data properties describe the relationships between individuals and data values. The proposed ontology contained the data properties such as ‘hasElevation,’ ‘hasPopulation,’ and the data properties defined in the GeoSPARQL standard. For instance, the property ‘hasElevation’ was of integer data, where ‘hasElevation’ was the property that described the elevation data of that place. All these properties and constraints enhanced ontology-driven knowledge management in the place ontology.

Comparison and Discussion

In this section, the developed ontology is compared with existing ontologies for the gazetteers created by the USGS, the Ordnance Survey, GeoNames, and OSM. Comparisons among these ontologies from multiple perspectives include developer, scope, dimension, language for implementation, and the URL of each ontology (Table

3.4). The URL hyperlink for each ontology can be opened using the ontology editor Protégé.

The ontology created by the USGS is designed to support the linking of cross-thematic datasets used in the national map (Regalia et al., 2018). The ontology of place names supports gazetteer information based on data from the GNIS, a geographic name repository for the USA. A classification of feature types accounts for a large portion of the GNIS ontology. One notable difference from the proposed place ontology is that each place-based class in the USGS ontology is disjointed from other parallel classes at the same level. In other words, every class is unique, and the classes do not share any instances. Because the proposed place ontology combined concepts from different sources, some concepts were subclasses of two different parent classes. For example, the class ‘Hospital’ was a subclass for both the classes ‘BusinessFeature’ and ‘SpotFeature.’ Hence, the proposed ontology did not define disjointed sets for classes. Instead, some equivalent classes were defined in the place ontology for different terms with the same semantic meaning but were extracted from different data sources.

The 50K Gazetteer is a place-name database designed in conjunction with a map published by the Ordnance Survey in the UK. The gazetteer contains all place names presented on the same scale of maps. As the Ordnance Survey has developed the ontology of the 50K Gazetteer to support the Ordnance-Survey-linked data, the ontology mainly focuses on gazetteer entries based on place names, feature types, and map references. The 50k Gazetteer ontology is a relatively small ontology with specific purposes. In contrast, the proposed place ontology showed more comprehensive place

knowledge containing gazetteer information, so this framework may help to populate place names with enriched information from a wide variety of data sources.

The GeoNames gazetteer is the most interlinked datahub of place names available online. The ontology was developed as a collaboration work, and it describes the elements of place descriptions defined in the GeoNames database. The geographic features can be interlinked based on the spatial relations applied. The GeoNames ontology enables the addition of geospatial semantic information to online databases and place ontologies. Likewise, the proposed place ontology was also developed to incorporate place information from various data sources. Moreover, the place ontology aimed to represent practical functions of places extracted from Twitter data and other spatial information involving places.

OSM provides a large dataset of tags to represent attributes of geospatial entities. The OSMonto has been developed based on the OSM database to describe the tags (Codescu et al., 2011). The OSM directly uses knowledge of places supplied by users (of, for example, social media) for the creation and enhancement of ontologies. Though OSM tags have the informal form of key=value, tags can also be enriched by using the relationships between external ontologies and the tag ontology. Due to the open structure of OSM, the OSMonto has considerable potential to enrich data for the OSM system. However, the unique system of terms and taxonomies of place names in the OSM makes it difficult to fully integrate information from other gazetteers. Like the OSM ontology, the proposed place ontology aimed to enrich information about places by aggregating public knowledge of places. However, extracted information from social media data was

designed to be added to other gazetteers in the place ontology to provide comprehensive knowledge about place names.

Table 3.4: A comparison of existing gazetteer ontologies

	Dev.	Scope	Dimension	Lang.
The place ontology¹²	Author of this study	Place names and related information in multiple gazetteers and Twitter data.	798 classes, 57 object properties, 11 data properties, and 952 logical axioms	OWL
GNIS ontology¹³	U.S. Geological Survey (USGS)	Geographic features based on GNIS data	71 classes, 7 object properties, 1 data property, and 84 logical axioms	RDF
GeoNames ontology¹⁴	Vatant & Wick	GeoNames toponyms and geospatial semantic information	11985741 features and about 182 mio rdf triples	RDF
50k Gazetteer Ontology¹⁵	Ordnance Survey	Place names in the 1:50 000 Scale Gazetteer presented by the OS	14 classes, 2 object properties, and 12 logical axioms	OWL
OSMonto¹⁶	Codescu, Horsinka, Kutz, Mossakowski, & Rau	Tags used in OpenStreetMap	466 classes, 29 object properties, 27 data properties, and 695 logical axioms	OWL

¹² https://github.com/doh30/Geo_Ontologies/blob/main/gaz_onto_v1.0.owl

¹³ https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/TOPO_Files.zip

¹⁴ https://www.geonames.org/ontology/ontology_v3.2.rdf,
https://www.geonames.org/ontology/mappings_v3.01.rdf

¹⁵ <https://www.ordnancesurvey.co.uk/legacy/ontologies/50KGazetteer.owl>

¹⁶ <https://raw.githubusercontent.com/doroam/planning-do-roam/master/Ontology/tags.owl>

In most cases, existing ontologies focus on the ontology components of their databases. Each gazetteer ontology could be mapped onto external database entries. However, this method enriches a gazetteer only partially or gradually. Therefore, there is a need to provide an overarching framework that combines the concepts and relations used in different gazetteers. In this sense, the proposed place ontology aimed to provide more comprehensive knowledge about place names and to support links to distributed gazetteer information.

Conclusion

This study explored diverse sources of information about place names and developed a place ontology. Taxonomic or descriptive information about place names, which were used in this study, are found in various types of sources. Regarding data sources, traditional gazetteers have been developed for official databases or for a specific purpose by an authority. Place names and relevant data are selectively recorded in these traditional gazetteers. More recently, non-traditional place-name repositories have appeared and include unofficial information about place names based on people's voluntary contributions to online maps (Keßler et al., 2009). Users are taking an active role in creating and updating place-name databases by voluntarily contributing information. In addition to the above two types of gazetteer, place information can be found in web resources. For instance, social media data were used in this study. Place descriptions extracted from social media data are often unstructured; therefore, they require data mining and natural language processing. Despite technical challenges, social media data have the advantage of providing up-to-date information, which is hard to find

in traditional gazetteers or non-traditional repositories of place names. The semantic representation across these heterogeneous data sources for place names has not yet been well established.

Overall, the current study advocates ontology-based semantic modeling of gazetteer information to store, manage, and infer domain knowledge. The major contributions of this research lie in two aspects. First, this study proposes an ontology-based knowledge representation and semantic reasoning framework to enrich information about place names. The development of the proposed ontology was conducted using the multiple data sources listed above and reusing some existing ontologies to represent topological properties from the OGC. This framework was then compared with existing gazetteer ontologies to compare the proposed approach with other approaches. Secondly, the proposed framework for creating a place ontology aligned the concepts and relations created by experts and non-expert users in diverse data sources, thus facilitating an open and shareable knowledge database for place information. This also implies that the place ontology can be enriched in a broad and deep manner by using external resources in further studies.

A summary of comparisons among existing gazetteer ontologies shows that these ontologies share the core concepts to represent place names, such as place types or spatial footprints. However, subsets of the respective ontologies need to be aligned for interoperability. Moreover, increasing amounts of place data are continuously available in the public domain on the internet, providing significant opportunities to enrich and update place databases. This study was, therefore, a research effort to provide technical solutions in responses to these problems and opportunities. It has considerable application

potential. For instance, the prototype of the proposed ontology shows the capability to retrieve place names and associated information using a wide range of search criteria. By doing so, the place ontology supports the discovery of knowledge about place names more than an independent gazetteer system does.

The proposed ontology is also amenable for customized developments and future research. For example, a place may be referred to by multiple names by different people or in different contexts. How to efficiently extract and organize these various references in the place database is worth more research effort. Further research should also be undertaken to investigate the conceptual links between places and the dynamic nature of places in more diverse aspects of human experiences. The constructed ontology is a starting framework that allows further developments for more in-depth insights into places.

CHAPTER 4
ENRICHED KNOWLEDGE ABOUT NAMED PLACES USING PLACE
DESCRIPTIONS EXTRACTED FROM TWITTER DATA

Introduction

While traditional gazetteers were used to design a conceptual place ontology, dynamic and evolving place characteristics were not captured. To overcome this weakness, the proposed ontology in this research uses social media data, specifically Twitter data in this case study, to explore how people describe place names based on their experiences and knowledge about those places in this dissertation. The assumption was that place names can be referenced using the extracted concepts from user-generated place descriptions. In Chapter 3, Twitter data, including a list of schools, colleges, and their building names (i.e., an education system in the study area), were used for a pilot study. As follow-up research, a set of place names recorded in the current gazetteers in the target study area (i.e., Athens, GA, USA) was used to examine subconcepts of the ‘function’ concepts that could reference place names and associated information in Chapter 4.

In addition to creating subconcepts of the ‘function’ concepts, this research attempted to answer how place type similarities can be measured using the extracted place-related concepts described by Twitter users to better understand place ontologies. The schemes of place type used in gazetteers represent a classification system for individual places. In gazetteers, place classifications are crucial elements for search of

places (e.g., schools in Clarke County, GA, USA). However, different typing schemes exist across gazetteers, which might cause difficulties in aligning gazetteers at both structural and semantic levels. Place classifications are used to describe places in different contexts, such as nature or urban spaces (e.g., mountains, streams, and population-centric places, such as cities). Place classifications are notably different between studies of nature and those of urban environments. However, the definitions of place categories are often ambiguous within urban environments because uses for places vary according to the needs of people. Therefore, this second pilot study proposed an approach to measuring similarities between place classifications based on functional signatures extracted from web-harvested place descriptions. In this study, functional signatures were related to people's activities regarding three factors: place affordances, events, and key-descriptors. The proposed approach was tested through a case study by matching place types used by local gazetteers using Twitter data. Results of this study are discussed for potential future improvements of the place ontology.

Place categories in gazetteers are typically not standardized. They are inherently disarranged and continually evolving according to user needs (Hastings, 2008; Janowicz & Keßler, 2008). The GNIS and GeoNames are excellent examples of how one typing scheme system is adopted from another. Both the GNIS and GeoNames adopted the place classifications developed by the U.S. NGA, which contains various place types for place names across the world. However, the GNIS was developed only for U.S. domestic place names, resulting in fewer place types and a coarser classification scheme than in the NGA. GeoNames includes place names from various countries using a variety of data

sources. Accordingly, GeoNames adopted a basic place-classification model from the NGA and introduced some additional classifications.

Another example of a typing scheme is the feature-type thesaurus (FTT), which was developed for the ADL Gazetteer (Hill, 2002). The FTT contains place types at multiple levels of a hierarchy, which is markedly different from the previous examples in terms of data and structure. Non-traditional data sources, such as Google Places, provide place categories that focus on urbanized areas, such as supermarkets, restaurants, car dealerships, and real estate agencies. This is because Google Places aims to provide users with information about local businesses, and it uses collaborative knowledge about places gathered from users. Considering the different aims of data sources, it might be challenging to establish a single scheme of place categories for all gazetteers. However, the similarity between place types in the same or different classification schemes can be measured to enhance current gazetteers by comparing thematic attributes extracted from web-harvested place descriptions.

Matching between place types should be built on thematic similarity, considering its original perspective. However, not every concept used in each system is clearly comparable among different gazetteers based on uniform categories. For example, POI, a popular place type used in gazetteers, is ambiguous in terms of its definition and functionality. Any place entity can be classified as both a POI and another place type. Likewise, places can be referred to by multiple place types. For example, the Georgia Museum of Art is categorized as 'Building' in the GNIS but grouped under 'Museum' in Google Places. These examples illustrate that each place category does not reflect all the properties of a place. More importantly, some place types share similar properties.

In general, place types could relate to each other in various ways based on perspective. For example, some place types are related to each other based on their spatial relationships which are rarely described in existing gazetteers (e.g., spatially, a university contains several buildings). In addition, taxonomies with *is-a* relationship are widely used in different typing schemes. Therefore, if a place is categorized at a lower level, that place is also considered an instance of the upper level of the classifications. For example, ‘Building’ is a parent concept of ‘Courthouse’ in the FTT; therefore, place entities classified as Courthouse are also classified as Building in the gazetteer. However, Building and Courthouse are defined at the same level of the typing scheme used by GeoNames, so there are separate data sets for each place type. Consequently, different schemes of taxonomies used for place classification may cause inconsistent search results between gazetteers. The approach suggested in this study can be used to assess similarities between place types regardless of the hierarchies used by different data sources, such as traditional and non-traditional place databases.

This research contributes to the discourse on digital gazetteers and GIR by comparing place types using thematic similarities. Descriptions of places online were used to extract people’s activities, which can be specified by three factors, to describe each place type: (1) place affordances, (2) events, and (3) key-descriptors. This approach included natural language processing (NLP), latent Dirichlet allocation (LDA) topic modeling (Blei, et al., 2003), and cosine similarity measurement. The three factors were then combined to determine the overall similarity scores for the comparison between place types.

This study utilized Twitter data from a study area, namely Athens, Georgia. However, the proposed approach can be applied to any user-generated content, such as blogs and articles, that describes places. The findings with high similarity scores can be summarized into three cases:

- (1) place types representing population-centric places, such as administrative regions, and classified as Populated Place;
- (2) place types describing the relation of spatial containment, such as POI and Populated Place; and
- (3) place types used for similar activities, such as weekend activities and relaxing, such as Park and Stream.

Using the overall similarity score alone was insufficient to align different place classifications because of the biased data source (e.g., user groups and topic frequency). However, the observations from the case study showed the possible relations between place classifications by using people's knowledge of places that can be added to the existing gazetteer alignment strategy.

The remainder of this chapter is organized as follows. Section 4.2 provides an overview of the related work, and Section 4.3 describes the proposed functional signature conceptual framework and the computational flowchart to assess place-type similarities, followed by a case study using the proposed method in Section 4.4. The results and major findings of the case study are discussed in Section 4.5. The chapter concludes with a summary and a discussion of future research directions in Section 4.6.

Related Work

Digital gazetteers are dictionaries of place names widely used in GIR (Goodchild & Hill, 2008). In gazetteers, places are identified based on several attributes, such as given names, classifications, and spatial footprints (Hill, 2009). Among place references used in gazetteers, a list of place classifications, such as parks, streams, neighborhoods, and libraries, is often labeled by the creators. Such place classifications are relatively subjective in describing places, incorporating human cognition and language (Mark & Turk, 2003). Moreover, these schemes are continually evolving according to needs (Hastings, 2008; Janowicz & Keßler, 2008).

For example, illustrating a hierarchical scheme of place classifications, Hill (2002) proposed the FTT developed in the ADL project. Subsequently, some studies have applied FTT as a place-type classification scheme for developing their place-type categories. Janowicz and Keßler (2008) proposed a feature-type ontology utilizing the FTT scheme to improve the interoperability and reasoning capability of gazetteers. The authors demonstrated that a feature-type ontology supports extended query functionality, which addresses the relationships between feature types.

Manguinhas, Martins, and Borbinha (2008) proposed a geotemporal gazetteer web service to integrate data from multiple sources, such as gazetteers of official toponymic authorities or public online sources with gazetteer data. Regarding the scheme of place classifications, the authors utilized the FTT with the classification schemes from the ECAI Time Period Directory and GeoNames to facilitate data integration from external sources.

Using an instance-matching approach, Brauner, Casanova, and Milidiú (2007) proposed an instance-based mapping rate between distinct FTTs by pre-processing common instances from two gazetteers, namely the GEOnet Names Server and the ADL Gazetteer. Several examples show that a mapping mechanism among different typing schemes is essential to enhancing the capability to answer queries about place information using multiple gazetteers.

Place can be described as a location that is filled with meanings created by human experiences (Tuan, 1977). Hence, place is difficult to describe objectively or define comprehensively (Winter, Kuhn, & Krüger, 2009; Goodchild, 2011). The proposed approach examines human interactions associated with different place types by focusing on people's activities. Therefore, a few studies were reviewed that focused on the thematic perspective of places used to define relationships between places based on shared common themes. For example, Adams (2015) and Adams, McKenzie, and Gahegan (2015) proposed an interactive thematic map search engine designed to search places based on extracted thematic topics (e.g., civil war) from web documents using LDA-based approaches.

Other research focused on historical records for places. Exploring narrative documents to detect historical events and places (e.g., January 7, 1859, in Wakulla County, FL, USA, as the day when the offices of Tax Assessor and Collector and Sheriff were combined) was studied as part of the Perseus Digital Library Project (Smith 2002). In addition, Mostern and Johnson (2008) proposed an approach to construct a historical event gazetteer using named places and historical events and visualized the links between such events and spatial changes. Therefore, a thematic perspective of places is often used

to explore similar places either alone or in combination with spatial and temporal perspectives.

Whereas traditional gazetteers created by authorities typically describe places using a formalized set of semantics, place descriptions generated by the general public with regard to named places reflect more varied aspects. A few examples of approaches that utilize place descriptions from web resources for exploring place semantics are discussed below. Purves, Edwardes, and Wood (2011) proposed a framework for gathering large collections of place descriptions from two different online communities, Geograph and Flickr. The authors addressed the problem that descriptions of places collected from different environments may vary in terms of sharing spatially referenced photographs with the associated content.

Kim, Vasardani, and Winter (2017) explored how place descriptions convey human spatial knowledge beyond GISs. Therefore, the same place could be described in multiple ways based on various place perspectives. The authors proposed a graph-based matching method for integrating spatial information extracted from various descriptions using string, linguistic, and spatial similarities. The approaches to finding corresponding places rely on spatial semantics matching with other types of similarity. The proposed approach to matching place categories combined place affordances and event and key-descriptors similarities. Because the proposed method focused on people's descriptions of their experiences in different places, the similarity of spatial relations was not used in the matching process. Still, it was described in the results of applying the approach.

Proposed Approach

The dissertation proposes to assess the similarity between two place types based on the designed functional signature. The concept of signature has been widely used in the literature. For research on places, the concepts of spatial, temporal, and thematic signatures have been developed and applied to capture spatial, temporal, and non-spatiotemporal characteristics of places and often used to classify places into place types (Adams & Janowicz, 2015; McKenzie et al., 2015; Zhu et al., 2016). In this dissertation, the concept of functional signature is proposed to capture unique functional characteristics of a place type. The functional signature consists of three component signature factors, including the Place Affordance, Events, and Key-descriptors.

Following the original definition of affordance, which is about opportunities that an environment has to offer (Gibson, 1979), the concept in spatial research is also approached from the individual perspective, and it is associated with the relationship between places and the action capabilities of individuals (Gibson, 1979; Raymond, Kyttä, & Stedman, 2017). Subscribing to this tradition, the place affordance of a place type in this study is represented by human activities at places of the type. The second component factor, Events, refers to the set of events that have taken place at places of the respective place type. In comparison, Place Affordance captures the functional characteristic of a place type from the perspective of individuals, while Events captures the functional characteristics from the perspectives of planners and practitioners. The third component factor, Key-descriptors of a place type, refers to the major terms that are frequently mentioned by people to describe the type of places. Examples are School, Food, Downtown, Park, Garden, etc.

Each of the three signature characteristics is presented as a vector. For any two place-types a and b , the three component signatures for each type are extracted from real-world data, for instance, the social media data. Then signatures of the pair of place types are compared to estimate the similarity between them from the perspective of each component. Finally, the three similarities from the component factors are summarized to give the final similarity measure between the two place-types.

The computational workflow consists of three steps: NLP, LDA topic modeling, and similarity measurement. Inputs are place descriptions that are collected from web documents or social media. Twitter data were used in a pilot study. In the first step, NLP is used to preprocess web-harvested place descriptions (Abilhoa and De Castro, 2014). After collecting place descriptions for a target area, keyword-based extraction is used for indexing and classifying the texts according to place categories. Keywords for detecting relevant texts are a set of place names for each place type sourced from multiple gazetteers.

A collection of place descriptions for each place type are seen as a document, $d_i = \{t_1^i, t_2^i, \dots, t_N^i\}$, where the document d_i consists of place descriptions containing the contents of place names for the i^{th} place type. Note that t_j is a collection of place descriptions associated with a specific place name.

In the second step, the LDA topic modeling is applied to identify the three component factors of the functional signature. LDA is a generative probabilistic model to automatically derive sets of words from a document to form latent topics (Blei, Ng, & Jordan, 2003). The input of the topic modeling model is the document for a specific place type. The document has a collection of descriptions for all places in the place type. It is

represented as a mixture of latent topics in a topic model, where each topic is characterized by a set of words. LDA is conditioned on three parameters: document–topic distribution α , topic–word distribution β , and topic number K . For a data set of tweets, the parameters are typically set as $\alpha = 0.1$, $\beta = 0.05$, and $K = 30$, and multiple parameters were tested to determine if the results showed any differences. The topic coherence score can be used to determine the quality of the learned topics. For each document d_i , topics are labeled for the three factors: place affordances, events, and key-descriptors.

In the last step, similarity scores among place categories are calculated for the three factors using a similarity measurement. For example, cosine similarity is a vector comparison measure widely used in information retrieval (IR), NLP, and text mining (Murphy, 2012; Li & Han, 2013). The cosine similarity of two n-dimensional vectors a and b are calculated using Equation (1), where a and b are two vectors for comparison, and a_i and b_i were components of vectors a and b , respectively. It is equal to the cosine of the angle between a and b . The output is a value between 0 and 1. A higher value means higher similarity between a and b .

Place affordances, events, and key-descriptors are used independently to measure the similarity among the various place categories. However, the similarity among some place affordances and key-descriptors may increase simply due to a recent event. Therefore, the final similarity is obtained as a weighted sum of the three component similarity measures, as represented in Equation (2). Here, S_O is the final similarity score; S_A , S_E , and S_K are three component similarity scores for affordability, events, and key-descriptors; and w_A, w_E, w_K are three respective weights that are customizable based on

context-contingent considerations. In the case study, for example, equal weights are taken, which means the three weights are 1/3 each.

$$\text{Similarity}(a, b) = \cos(\theta) = \frac{\sum_{i=1}^n a_i \times b_i}{\sqrt{\sum_{i=1}^n (a_i)^2} \times \sqrt{\sum_{i=1}^n (b_i)^2}} \quad (1)$$

$$S_O = w_A \times S_A + w_E \times S_E + w_K \times S_K \quad (2)$$

Case Study

Study Area and Data

A pilot study using the proposed approach was conducted on a combined set of place-type classifications from three gazetteers for a target study area (i.e., Athens, GA, USA). The three gazetteers were the GNIS, GeoNames, and Google Places. The GNIS, developed by the USGS, is the official repository of U.S. domestic geographical name data. Place types were defined as feature classes containing all feature types for place names, except for road and highway names. The GNIS feature classes apply the second level of classifications created by the NGA.

GeoNames is a global geographic database that includes over 11 million place names worldwide, and it can be accessed via its web services. GeoNames primarily uses the same feature code as the NGA, with a two-level hierarchy method. Together with these feature codes, GeoNames uses additional codes if necessary.

Google Places' service provides a place-searching capability with a list of place categories and detailed information about a specific location. Place types used in Google Places include various types of business, such as clothing stores and home goods stores, which are not covered by the other two gazetteers. For the current study, these store data sets were combined under 'Business' to collect sufficient relevant Twitter data for the

analysis. In addition, the locality and postal code of the area were combined under ‘Region’ for the same reason. As noted, specific types combined under Business and Region had finer level of categories than others.

Many place types are often shared among different sources. However, a place could be referenced with different categories by various creators. For example, the Athens Regional Library is categorized as Building in the GNIS but as Library in GeoNames. Thus, a place may be duplicated across place types. Table 4.1 summarizes the statistical features of the place categories in the target study area using the three gazetteers.

Table 4.1: Summary of statistical features of place categories in Athens, GA, USA

Type	# of Features	Mean Dist	Max Dist	Std Dev	Entropy
ADM2	1	NA	NA	NA	5.56
Building	80	0.03	0.19	0.03	2150.96
Business	1372	0.07	0.28	0.04	31056.33
Church	185	0.07	0.25	0.04	2722.72
Civil	6	0.08	0.2	0.05	25.06
Hospital	33	0.03	0.15	0.03	535.9
Library	23	0.05	0.18	0.04	434.92
Neighborhood	18	0.08	0.17	0.05	67.49
Park	87	0.05	0.22	0.03	1479.89
Point of Interest	579	0.1	0.29	0.05	5298.91
Political	38	0.1	0.26	0.05	128.9
Populated Place	125	0.08	0.23	0.04	726.82
Region	10	0.09	0.24	0.05	86.02
School	175	0.06	0.27	0.04	2752.18
Stadium	1	NA	NA	NA	11.87
Stream	16	0.11	0.28	0.06	27.47

The total number of places per place category varied, partly because of actual differences and partly because places were classified differently in the gazetteers. Spatial patterns were identified based on the minimum and maximum distances between the places for each place type, along with its standard deviation. The entropy was calculated to measure the spatial relations between place categories. The entropy of the i^{th} place type is defined in Equation (3).

$$E_i = \sum_{i=1}^n \frac{n_i}{N} \log \left(\frac{n_i}{N} \right) \quad (3)$$

where N is the total number of place types in the area and n_i is the number of nearest instances from the i^{th} place type (Zhu, Hu, Janowicz, & McKenzie, 2016). A larger entropy value indicates a greater variety of places around the target place type.

Twitter data were used as the source of place descriptions for this case study. The data set was collected using the bounding box of the target study area between December 2017 and January 2019. The Twitter data collected were then filtered to identify texts that described places only. Place names were collected from the three gazetteers (i.e., GNIS, GeoNames, and Google Places) by place type and used as a set of keywords to extract tweets. This dissertation assumed that the combined texts categorized by place type contained several topics that could be used as place-type references. Therefore, the data sets were processed using a topic-modeling method to identify three factors (i.e., place affordances, events, and key-descriptors) based on the calculated similarity scores. Figure 4.1 shows the numbers of texts collected in the study area for a sample of place types. Athens is a college town in the USA, so many tweets describe topics related to the schools and colleges.

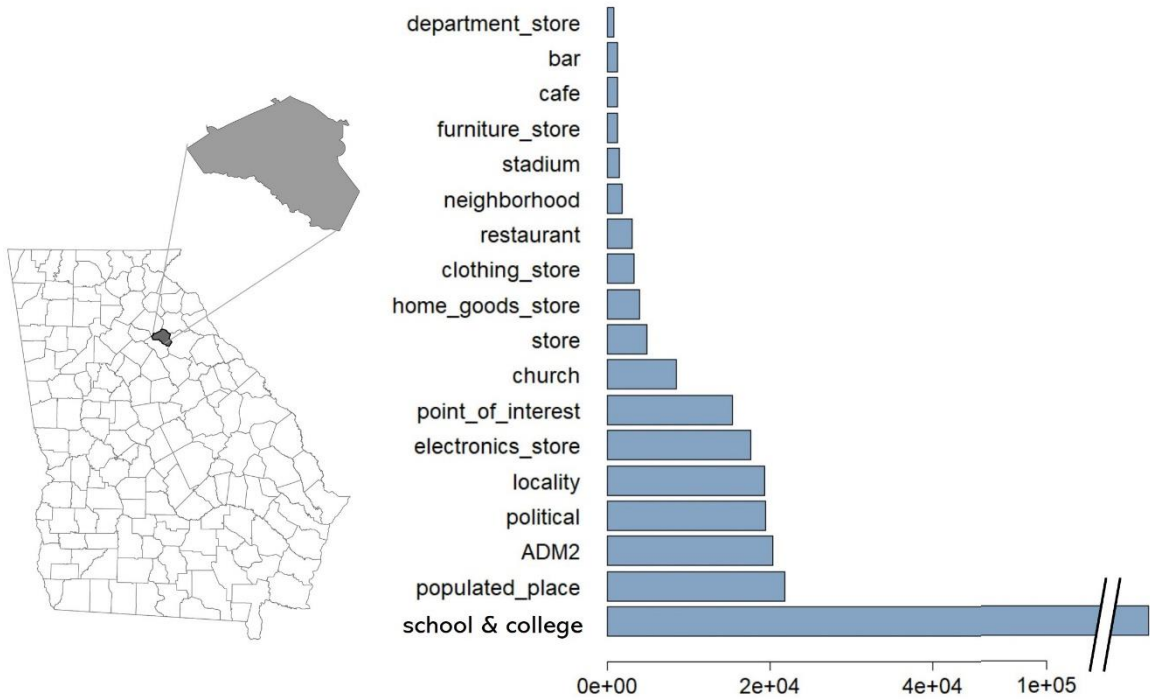


Figure 4.1: The number of tweets collected for each place type in the target study area (Athens, GA, USA, on the left-hand side)

Similarity Measurement

It was assumed that some properties were shared between place types based on people's activities that can be used to measure similarities between place types. Place types identified by more than 50 relevant tweets in the study area were included for the analysis. In total, 91 place types were detected in more than one tweet. Among them, 14 place types had enough tweets and were selected for the similarity measurement.

Using LDA, the top 20 key terms for 50 topics were calculated for 14 place categories. LDA was performed for 2000 iterations using the Gibbs sampling algorithm for each data set. Then, latent topics were manually labeled for the three factors: place

affordances, events, and key-descriptors. Table 4.2 shows the similar textual patterns in topics identified for two different place types: ADM2 (second-order administrative division) and Stadium. Some place categories were described by specific features. For example, schools and colleges were frequently mentioned with building names, school activities, and sports events. Streams were often mentioned with other place types, such as rivers, creeks, dams, and parks.

Table 4.2: Examples of word sets forming topics for the place categories, ADM2 and Stadium

Place type	Topic	A collection of words
ADM2	Football game	go dawgs, dawgs on top, Georgia football, game Sanford stadium, ...
	School	university, Sanford stadium, college students, school Athens...
	Job marketing	job, GA job, job Athens, hiring, recommend job, apply...
	Weather reporting	weather summary, forecasts UGA Sanford, sunrise ...
	Weekend activity	weekend, downtown weekend, getaway, camping, ...
Stadium	Football game	go dawgs, Georgia football, game Stanford, AU vs UGA, ...
	School	university, university Georgia, UGA, students, ...
	Weather reporting	weather hours, raining UGA Sanford, day forecasts, ...

Among the specific features used to describe place categories, events were more likely to be regional features. In general, extracted event names reflected locals' interests, such as school, crime, and sports games. Topics related to football games frequently

appeared and formed a large proportion of the Twitter data in the study area. This resulted in an increase in associated activities and key-descriptors used for football games for a specific time period. Therefore, similarity measurement between place types necessitated a comprehensive understanding of people’s activities, considering the three factors simultaneously.

Similarity scores between place types were measured using cosine similarity. First, cosine similarity was calculated for each of the three factors: place affordances, events, and key-descriptors. Figure 4.2 shows the low-to-high similarity scores among the place types for each factor. Similarity scores for the events factor tended to have higher

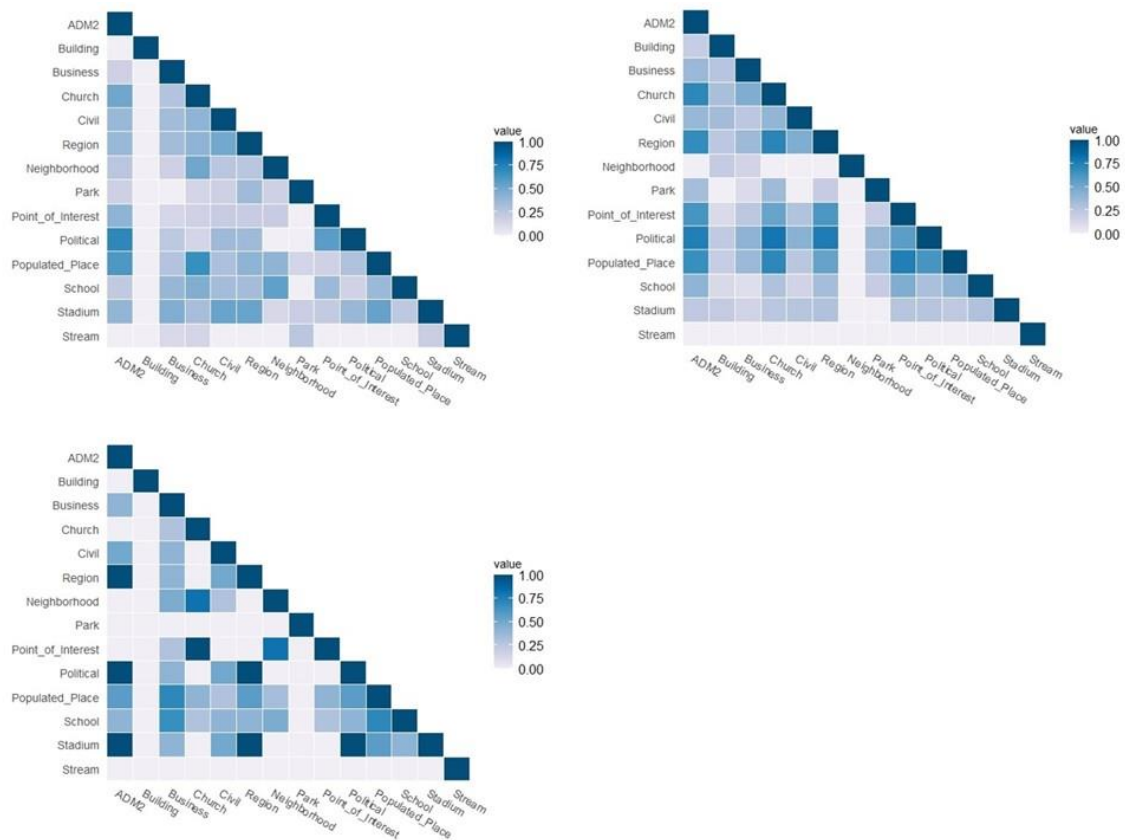


Figure 4.2: Cosine similarities between place categories by individual factors: place affordances, key-descriptors (top, left to right), and events (bottom)

values than the other two factors, because the total number of events identified from the Twitter data was small. Subsequently, the overall similarity scores were calculated using Equation (2). Table 4.3 shows the overall similarity scores for place types with individual scores based on the three factors. It is evident that the highest similarity scores were found between population-centric place classifications, such as ADM2, Political, Region, and Populated Place. In addition, based on people’s activities extracted from the Twitter data, patterns were discovered between two place categories based on the spatial containment relationship, such as Church and Populated Place, and two distinct place categories, such as Business and School. More details are provided in Section 4.5.

Table 4.3: Overall similarity scores for distinct place categories, with the individual similarity scores for three factors

Type A	Type B	Activity Similarity	Event Similarity	Key-descriptors	Overall Similarity
ADM2	Political	0.707107	1	0.755929	0.821012
Region	Political	0.353553	1	0.771517	0.708357
ADM2	Region	0.375	1	0.680414	0.685138
ADM2	Populated Place	0.612372	0.57735	0.666667	0.618796
Region	Stadium	0.534522	1	0.288675	0.607733
Church	Populated Place	0.666667	0.408248	0.707107	0.594007
Church	Point of Interest	0.166667	1	0.534522	0.567063
Political	Stadium	0.377964	1	0.267261	0.548409
ADM2	Stadium	0.400892	1	0.235702	0.545531
Region	Populated Place	0.408248	0.57735	0.544331	0.509977
Political	Populated Place	0.288675	0.57735	0.629941	0.498655
Populated Place	School	0.365148	0.707107	0.408248	0.493501
Civil	Region	0.5	0.5	0.46291	0.487637

Populated Place	Stadium	0.545545	0.57735	0.235702	0.452866
Business	Populated Place	0.280056	0.707107	0.363696	0.450286
Point of Interest	Populated Place	0.166667	0.408248	0.755929	0.443615
Church	Neighborhood	0.51031	0.816497	0	0.442269
Civil	Stadium	0.534522	0.5	0.267261	0.433928
Civil	Political	0.353553	0.5	0.428571	0.427375
ADM2	Civil	0.375	0.5	0.377964	0.417655
ADM2	Church	0.51031	0	0.707107	0.405806
Point of Interest	Political	0.57735	0	0.571429	0.382926
Business	School	0.383482	0.666667	0.089087	0.379745
Church	Region	0.408248	0	0.721688	0.376645
Point of Interest	School	0.365148	0.288675	0.46291	0.372245
Business	Region	0.342997	0.408248	0.356348	0.369198
Region	School	0.33541	0.408248	0.333333	0.358997
Business	Political	0.242536	0.408248	0.412393	0.354392
ADM2	School	0.223607	0.408248	0.408248	0.346701
ADM2	Point of Interest	0.408248	0	0.629941	0.346063
Church	School	0.456435	0.288675	0.288675	0.344595
Business	Church	0.280056	0.288675	0.46291	0.34388
Neighborhood	School	0.559017	0.471405	0	0.343474
Business	Stadium	0.458349	0.408248	0.154303	0.3403
Neighborhood	Point of Interest	0.204124	0.816497	0	0.340207
Business	Civil	0.342997	0.408248	0.247436	0.332894
Church	Political	0.144338	0	0.801784	0.315374
ADM2	Business	0.171499	0.408248	0.363696	0.314481
School	Stadium	0.239046	0.408248	0.288675	0.31199
Civil	School	0.33541	0.408248	0.154303	0.299321
Political	School	0.158114	0.408248	0.308607	0.291656
Civil	Populated Place	0.306186	0.288675	0.251976	0.282279
Region	Point of Interest	0.204124	0	0.617213	0.273779
Church	Civil	0.408248	0	0.400892	0.269713
Business	Neighborhood	0.171499	0.471405	0.154303	0.265735
Neighborhood	Populated Place	0.408248	0.333333	0	0.247194

Business	Point of Interest	0.140028	0.288675	0.247436	0.22538
Church	Stadium	0.327327	0	0.25	0.192442
Region	Park	0.353553	0	0.204124	0.185893
Civil	Neighborhood	0.25	0.288675	0	0.179558
ADM2	Park	0.176777	0	0.333333	0.170037
Church	Park	0.144338	0	0.353553	0.165964
Civil	Point of Interest	0.204124	0	0.285714	0.163279
Point of Interest	Stadium	0.218218	0	0.267261	0.161826
Park	Populated Place	0.144338	0	0.333333	0.159224
Park	Political	0	0	0.377964	0.125988
Building	Civil	0	0	0.341882	0.113961
Building	Church	0	0	0.319801	0.1066
Building	Business	0	0	0.263181	0.087727
Park	Stream	0.25	0	0	0.083333
Region	Neighborhood	0.25	0	0	0.083333
ADM2	Neighborhood	0.25	0	0	0.083333
Building	Region	0	0	0.246183	0.082061
Building	Political	0	0	0.227921	0.075974
Building	Stadium	0	0	0.213201	0.071067
Building	Neighborhood	0	0	0.213201	0.071067
Park	School	0	0	0.204124	0.068041
Building	Populated Place	0	0	0.201008	0.067003
ADM2	Building	0	0	0.201008	0.067003
Stadium	Stream	0.188982	0	0	0.062994
Park	Stadium	0.188982	0	0	0.062994
Park	Point of Interest	0	0	0.188982	0.062994
Neighborhood	Park	0.176777	0	0	0.058926
Civil	Park	0.176777	0	0	0.058926
Church	Stream	0.144338	0	0	0.048113
Neighborhood	Stadium	0.133631	0	0	0.044544
Building	School	0	0	0.123091	0.04103
Business	Stream	0.121268	0	0	0.040423
Building	Point of Interest	0	0	0.113961	0.037987
Business	Park	0	0	0.109109	0.03637

Results and Discussion

Place Affordances, Events, and Key-descriptors

The results for the three factors (i.e., place affordances, events, and key-descriptors) reflected the characteristics of the study area, and these can be expanded for other case study areas in future studies. Place affordances included different kinds of activity, such as economic, social, fun, and service activities. Among 37 place affordances for 14 place types, ‘hiring,’ ‘weather reporting,’ and ‘game activity’ frequently appeared for different types of place (Figure 4.3). For example, hiring might not be a common place reference in existing place-name databases. However, place types for hiring indicated that these places were where people worked, such as business or service types of place. Therefore, places were categorized for hiring and linked to associated information.

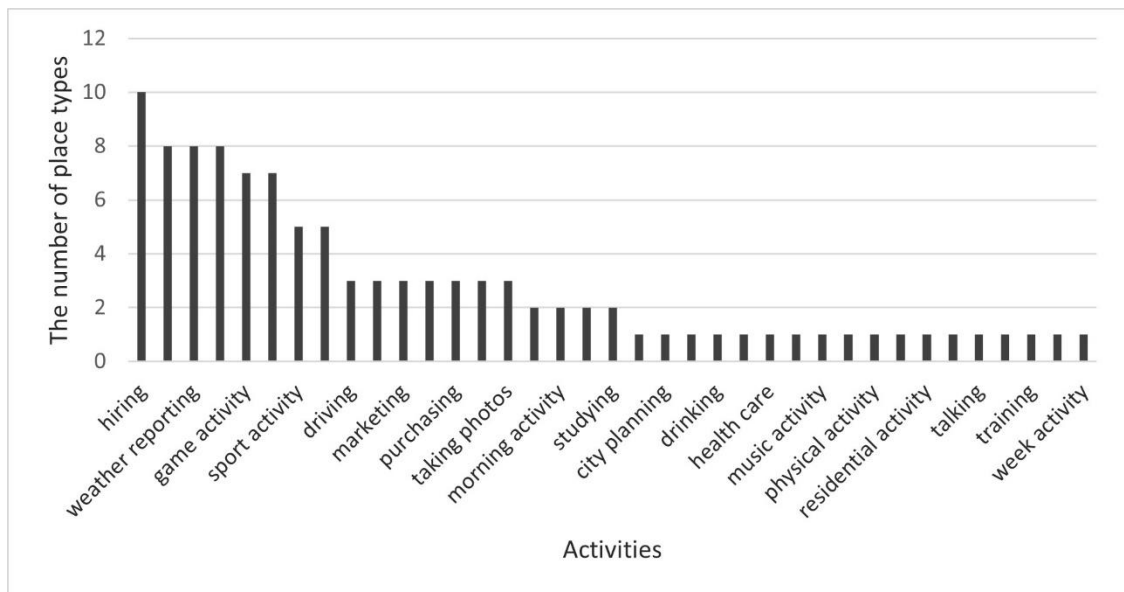


Figure 4.3: Multiplicity of place types vs activity types

Events extracted from Twitter data in the study area include sport, social, and political events (Figure 4.4). Because football games are of the most significant interest to students and locals, the SEC Championship and Rose Bowl frequently appeared for different place types. However, these topics are not only related to places where football players play games, such as a stadium, but also related to places where people watch games, experience tailgating, or associate them with the name of an administrative region (e.g., football game in Athens). Therefore, the event concept could be used to find place names and associated information either independently or together with other place-related concepts (e.g., to find place names using place type and event instances).

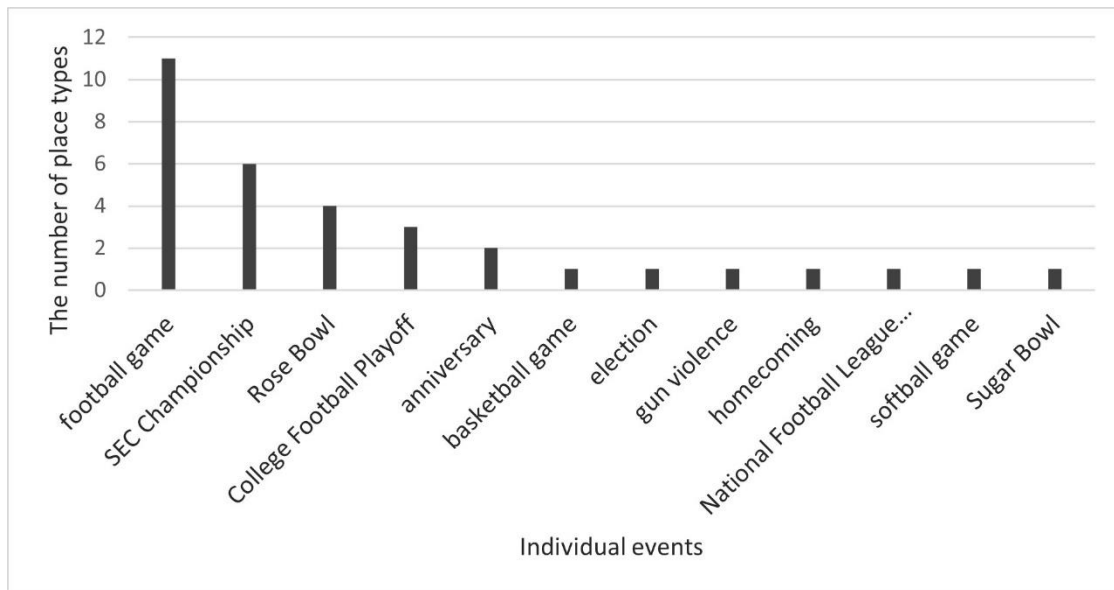


Figure 4.4: Multiplicity of place types vs event types

Key-descriptors extracted for different place types included specific types of business that might be grouped into other place types in existing place-name databases (e.g., waffle place into restaurant). Key-descriptors were discovered mainly from places that have relatively large areas. First, places that have several buildings or facilities are good examples (e.g., schools & colleges). Secondly, administrative regions are usually referenced for any places in that area. Finally, when man-made features exist near natural environments, they are frequently mentioned together (e.g., dam and river). Therefore, place types that co-appeared more frequently together were more related to each other.

Future studies will incorporate the three factors extracted into the proposed place ontology as subconcepts of functions concepts. In addition, additional descriptive information about places extracted through other case studies will be also considered to describe functional characteristics for expanded knowledge about places and improvements in IR.

High similarity between place groups of population-centric places

The highest similarity scores were observed between population-centric place types, such as ADM2, Political, Region, Populated Place, and Civil. Although the concepts and definitions are all different, they provide similar functionality with slightly different application areas. For example, Clarke County is an instance of ADM2 and Civil in GeoNames and the GNIS gazetteer, respectively. In these two gazetteers, Civil includes a broad range of administrative divisions, such as borough, county, incorporated place, and township. ADM2 covers the second level of administrative division, which is a county-level unit in the USA. The definition of Political used in Google Places is not

provided, but it covers county, city, and other administrative units. The number of places related to these five different concepts are listed in Table 4.4.

Table 4.4: The number of instances for selected place categories in Athens, GA, USA

	GNIS	GeoNames	Google Places
ADM2	NA	1	NA
Civil	6	NA	NA
Political	NA	NA	38
Populated Place	125	123	NA
Region	NA	NA	10

These similarity patterns can be used in addition to existing alignment methods for place classification based on the highest similarity scores among population-centric place types calculated using people’s activities extracted from Twitter data. The proposed approach for matching different place categories alone may have had lower discriminatory power. However, it was not influenced by other alignment techniques for place types that use spatial signatures (Zhu, Hu, Janowicz, and McKenzie, 2016) or an instance-based matching approach (Brauner, Casanova, and Milidiú, 2007). Instead, it revealed thematic patterns and people’s daily use of places.

High similarity between place types describing the spatial containment is-a and part-of relationships

In the place descriptions that people created, they tended to refer to the place in the larger area unit (e.g., the University of Georgia in Athens). This pattern was well

described in the similarity scores among some place types and population-centric place types. The typing schemes used in the three gazetteers (i.e., GNIS, GeoNames, and Google Places) do not describe the relationships of spatial containment between the types. The limited information on semantics for place classification results in significant challenges for developing ontologies. High similarity scores among place types that show spatial containment relations could help construct place-type ontologies; for example, high similarity scores between Populated Place and Church.

Other examples include place types, such as POI or Building, that are not defined clearly. They are the place categories used in existing gazetteers that share the same instances with other place types within or across gazetteers. For example, POI includes any place that people may consider interesting. It could encompass residential areas, schools, parks, restaurants, and churches. Therefore, the high similarity scores between POI and other place types such as Populated Place and School are expected. Figure 4.5 illustrates the relative positions of place instances for three place types, POI, Populated Place, and School, using multidimensional scaling (MDS). MDS represents the information contained in a distance matrix in order to visualize the similarity or dissimilarity among the investigated data sets. In the MDS transformed feature space, similar places will have shorter distances between each other than dissimilar places. The individual places of these three place types are relatively close and have significant overlapping positions.

Building is another example that shares common properties with other place types. In the GNIS and GeoNames, Building describes a distinct place type, not including other types such as Church, Hospital, or School. However, buildings are individual

structures that can be a part of different place types. For instance, the Georgia Center, which was built as a multipurpose construction serving as a hotel and a conference center, is classified under Building in the gazetteers but is, in fact, also part of a university (place type School). Therefore, individual buildings are often described with other place types (key-descriptors) in people’s descriptions of these places on social media.

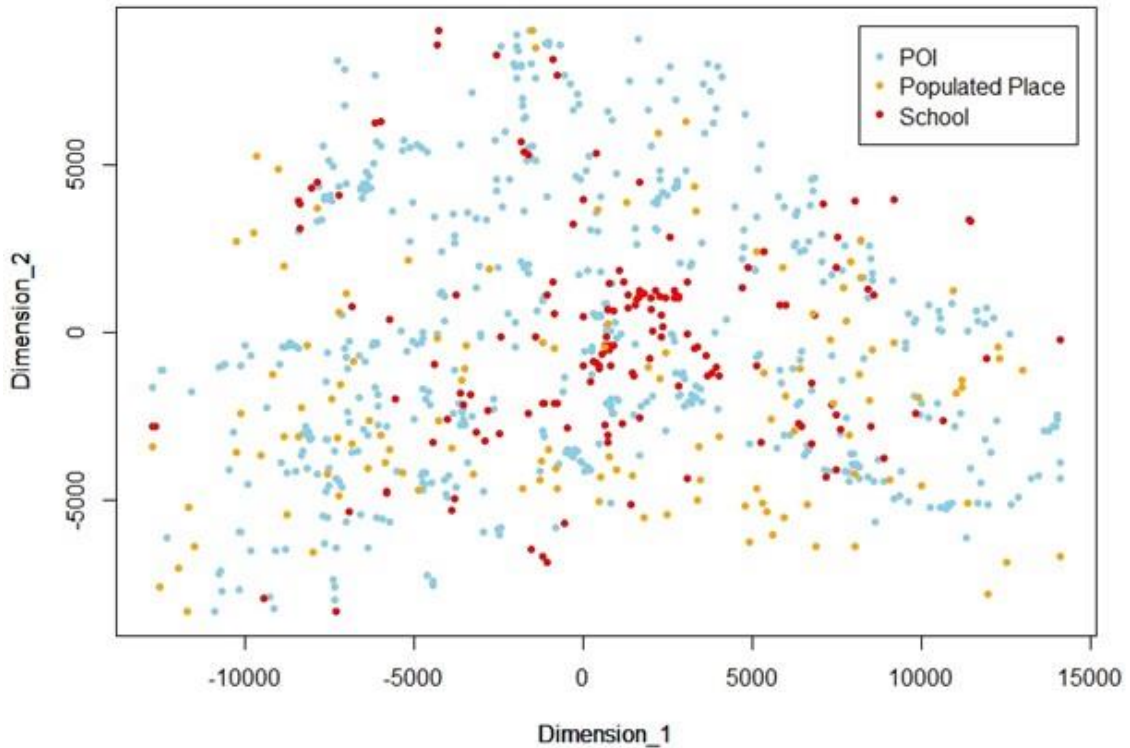


Figure 4.5: Multidimensional scaling (MDS) map for POI, Populated Place, and School

High similarity between two distinct place-type groups

Although some place types were not strongly related spatially, high similarity scores were identified between these place types according to the three factors based on

people's activities. For example, churches and neighborhoods in the study area were not close neighbors in a 2D representation (Figure 4.6). However, Church and Neighborhood were frequently referred to in similar types of activity, such as sport, tours, and weekend activities, in the local area.

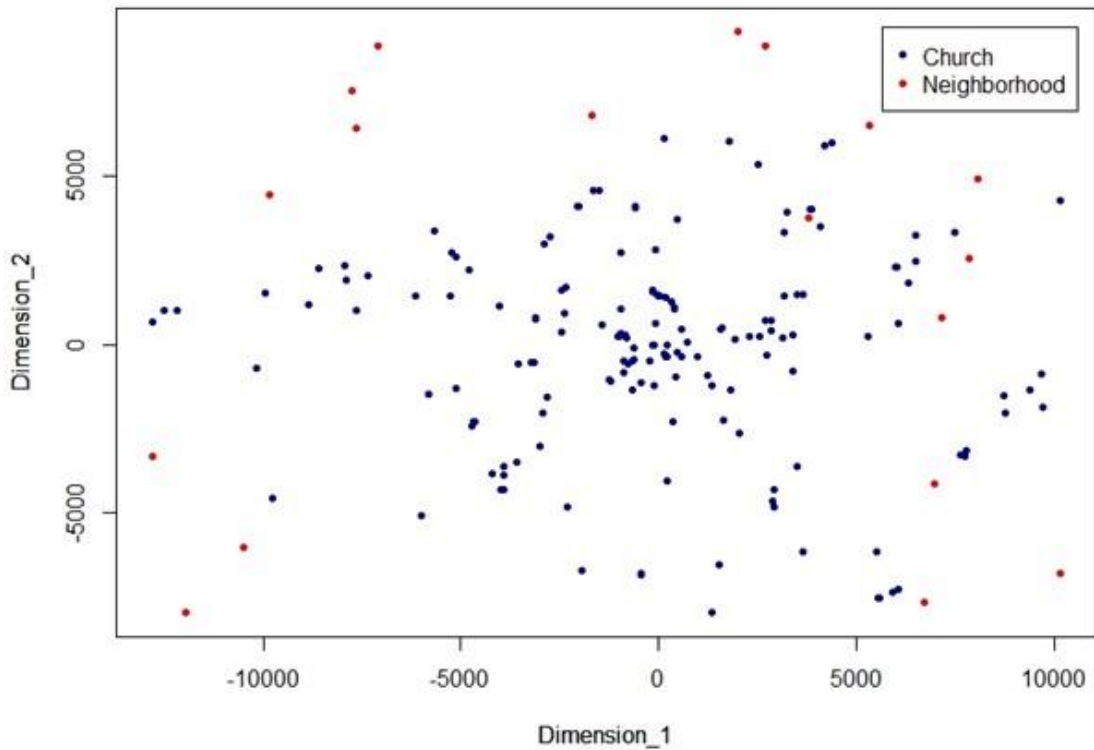


Figure 4.6: Multidimensional scaling (MDS) map for Church and Neighborhood

Another example is the high similarity scores identified between Business and School. These two types shared many everyday activities, such as entertainment, sport, political activities, and service activities. Key-descriptors were also observed for place types like gardens and schools. High similarity between two distinct place-type groups,

which are spatially separated from each other, show that those two groups are meaningful in human behaviors and place uses. However, this case is not sufficient for direct use in gazetteer alignment processing because of the lack of spatial semantics. Nevertheless, it can be used to search place names based on similar activities (thematic signature) in local gazetteers. This study highlights that there are pairs of place types that share common activities.

Finally, the results allow us to compare different place types across gazetteers or search local gazetteers based on people’s activities. A preliminary mapping based on the case study is illustrated in Figure 4.7.

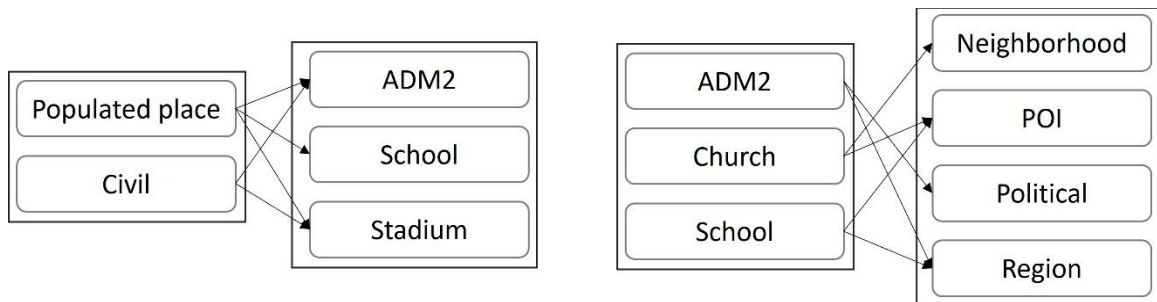


Figure 4.7: Place types matching between GNIS and GeoNames (left), and GeoNames and Google Places (right) based on the thematic characteristics

Conclusion

This research presents the approach for extracting place-related concepts and a similarity measurement among place categories using place descriptions created by Twitter users. The three factors, place affordances, events associated with place types, and key-descriptors, are not the common concepts to reference place names and places in

the existing place-name databases. These concepts can be harvested from web resources, such as news articles, blogs, and social media, to provide dynamic aspects of place uses that can be used as place references. These three concepts will be harvested for more areas and will be incorporated into the place ontology for GIR in future studies.

This chapter focused on measuring the similarities among place types to better understand a descriptive concept of place names. The similarity scores were calculated for three factors: place affordances, events associated with place types, and key-descriptors. The three respective values were combined to obtain the overall similarity score between place types. Such bottom-up approaches using user-generated descriptions are necessary for determining how people utilize places. People's experiences in places are not limited to the intended purposes of the places. To demonstrate the proposed approach in a case study, Twitter data were used as a source of place descriptions because of its data availability and because Twitter has recently become the most popular social media platform for people to share their thoughts and activities. However, the proposed approach is generally applicable to other types of online sources as well.

A noticeable difference in similarity scores for each factor among place types was not demonstrated in this study. In other words, place types with a high similarity score based on place affordances usually also showed a high similarity score for events. However, place affordances were more related to daily activities, while events could be seasonal. It is believed that some place affordances and key-descriptors observed for the place types could be affected by an event in a specific time frame (e.g., tailgating at a football game at the school). Thus, the overall similarity score is more powerful for comparing different place types than are the three individual similarity measures. Overall,

high similarity scores were identified among place types with overlapping functionalities, such as population-centric places, those representing spatial containment relations, or two distinct place-type groups. The results can be used to compare different place categories across gazetteers or search local gazetteers based on people's activities.

Compared with the existing methods for aligning place categories, the proposed approach had a strong ability to explain the relations between place types based on people's daily activities. This study has several limitations. First, in spite of the many benefits of using social media data, such data are usually biased and thus are not entirely random samples. For instance, prior studies found that social media data do not proportionally represented different age groups and different social groups. Second, only the place types that were mentioned in sufficient social media posts could be analyzed to measure similarities. Therefore, some place types may have been omitted due to sampling bias. Future studies will focus on different typing scheme alignments using the results of this research along with the existing matching strategies. Therefore, this research aims to better understand typing scheme to enrich gazetteers.

CHAPTER 5

CONCLUSION

This dissertation has focused on building a conceptual ontology for place names and their semantic information (e.g., place types of place names). As traditional gazetteers provide limited information about place names, there is a significant need to enrich existing gazetteers using various data sources for the purpose of GIR. Non-traditional gazetteers, as volunteer-led and open-source projects, provide unofficial and colloquial information about places. However, they are not easily interoperable with traditional gazetteers because of their different schemes. In addition to non-traditional gazetteers, such as Google Places and OSM databases, user-generated place descriptions from web-harvested resources provide more information about places related to people's activities. The goal of developing the ontology is to provide semantic information about place names for traditional and non-traditional gazetteers and user-generated place descriptions. Therefore, the developed place ontology can be used to find place information using the relationships between concepts adopted from multiple data sources to represent attributes of place names.

Three objectives were addressed in this dissertation. The first objective was to identify different schemes used in different data sources of place names and associated place information. In the data sources used in this dissertation (i.e., GNIS, GeoNames, TGN, Google Places, and Twitter data), place names are described using several place elements and relevant information. For example, place names, place types, and spatial

footprints are the core components of gazetteers. Statistical data provide additional information about places and help in understanding areas and in planning. Functions concepts were added for user-generated place descriptions. The subconcepts of functions were extracted from Twitter data in the study area (Chapter 3 [only for schools and colleges as a case study] and Chapter 4).

The second objective was to formalize the defined concepts and relations in order to represent place names and associated information about places. The proposed place ontology included six top-level concepts and the properties that represented the relations among those concepts. Most gazetteers describe spatial footprints by point locations using a coordinates system. However, places can also be described using spatial relations among their neighbors. The place ontology included GeoSPARQL, which is used for representing geometries and spatial relations of geospatial data. The place ontology implemented and tested the logical consistency in Protégé (Chapter 3). Because it was developed for multiple place-name databases, this ontology retrieved more comprehensive knowledge about places than the existing gazetteer ontologies.

Among the gazetteer components, place types are descriptive concepts to represent the purpose of places. Whereas some place references, such as place names and spatial footprints, can be used to navigate to a specific location, place types are categorical data. The third objective focused on matching similar place categories based on people's activities (Chapter 4). As a case study, Twitter data in the study area were collected and analyzed to extract three factors (i.e., place affordances, events, and key-descriptors). Among the three factors, place affordances and events factors are subconcepts of the functions concepts in the proposed place ontology. This will be tested

for more case studies and implemented in the ontology in future studies. In addition, the results of place-type similarities show the significant potential that the three factors can be used for gazetteer matching based on people's activities. This research highlighted that gazetteers can be enriched using multiple data sources and an ontological approach. To improve query results using the developed ontology, it should be extended for various data sources, including extensive areas with distinct named places.

This dissertation contributes to discourses on digital gazetteers and GIR for better understanding of how place information can be represented and retrieved using an ontological approach. This project considers multiple place-name data sources for collecting place names and associated information to provide comprehensive knowledge about named places. However, place information is not limited to traditional and non-traditional gazetteers and social media data. Place knowledge can be enriched through various other channels. For example, local news articles are the most up-to-date information sources. These are just a few examples that can be used to represent knowledge about named places and can be added to the ontology. More studies are needed to investigate the relationships among place-related concepts in future studies.

BIBLIOGRAPHY

- Abilhoa, W. D., & De Castro, L. N. (2014). A keyword extraction method from twitter messages represented as graphs. *Applied Mathematics and Computation* 240, 308-325.
- Adams, B. (2015). Finding similar places using the observation-to-generalization place model. *Journal of Geographical Systems* 17(2), 137-156.
- Adams, B. and Janowicz, K. (2015). Thematic signatures for cleansing and enriching place-related linked data, *International Journal of Geographical Information Science*, 29:4, 556-579.
- Adams, B., McKenzie, G., & Gahegan, M. (2015). Frankenplace: interactive thematic mapping for ad hoc exploratory search. In *Proceedings of the 24th international conference on world wide web*, 12-22.
- Ballatore, A. (2016). Prolegomena for an ontology of place. *Advancing geographic information science*, 91-103
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Borges, K. A., Davis, C. A., Laender, A. H., & Medeiros, C. B. (2011). Ontology-driven discovery of geospatial evidence in web pages. *Geoinformatica*, 15(4), 609-631.

- Borges, K. A., Laender, A. H., Medeiros, C. B., & Davis Jr, C. A. (2007). Discovering geographic locations in web pages using urban addresses. *Proceedings of the 4th ACM workshop on Geographical information retrieval*, 31-36.
- Brauner, D. F., Casanova, M. A., & Milidiú, R. L. (2007). Towards gazetteer integration through an instance-based thesauri mapping approach. *Advances in Geoinformatics*, 235-245. Springer, Berlin, Heidelberg.
- Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them?. *IEEE Intelligent Systems and their applications*, 14(1), 20-26.
- Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T., & Rau, R. (2011). Osmonto-an ontology of openstreetmap tags. *State of the map Europe (SOTM-EU) 2011*.
- Cresswell, T. (2014). *Place: an introduction*. John Wiley & Sons.
- Cruz, I. F., Sunna, W., & Chaudhry, A. (2004). Semi-automatic ontology alignment for geospatial data integration. *International Conference on Geographic Information Science*, 51-66. Springer, Berlin, Heidelberg.
- Fonseca, F. T., Egenhofer, M. J., Davis Jr, C. A., & Borges, K. A. (2000). Ontologies and knowledge sharing in urban GIS. *Computers, Environment and Urban Systems*, 24(3), 251-272.
- Fonseca, F., Egenhofer, M., Davis, C., & Câmara, G. (2002). Semantic granularity in ontology-driven geographic information systems. *Annals of mathematics and artificial intelligence*, 36(1-2), 121-151.

- Frank, A. U. (2001). Tiers of ontology and consistency constraints in geographical information systems. *International Journal of Geographical Information Science*, 15(7), 667-678.
- Gelernter, J., Ganesh, G., Krishnakumar, H., & Zhang, W. (2013). Automatic gazetteer enrichment with user-geocoded data. *Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*, 87-94.
- Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin Harcourt, Boston.
- Golledge, R. G. (2002). The nature of geographic knowledge. *Annals of the Association of American geographers*, 92(1), 1-14.
- Goodchild, M. F. (2004). The Alexandria digital library project. *D-Lib Magazine*, 10(5), 1-8.
- Goodchild, M. F. (2011). Formalizing place in geographic information systems. *Communities, neighborhoods, and health*, 21-33. Springer, New York, NY.
- Goodchild, M. F., & Hill, L. L. (2008). Introduction to digital gazetteer research. *International Journal of Geographical Information Science*, 22(10), 1039-1044.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), 199-220.
- Guarino, N. (Ed.). *Formal Ontology in Information Systems. (Proceedings of FOIS' 98, Trento, Italy, 6-8 June 1998)*, Amsterdam: IOS Press.
- Hastings, J. T. (2008). Automated conflation of digital gazetteer data. *International Journal of Geographical Information Science* 22(10), 1109-1127.

- Hill, L. L. (2000). Core elements of digital gazetteers: placenames, categories, and footprints. *International Conference on Theory and Practice of Digital Libraries*, 280-290. Springer, Berlin, Heidelberg.
- Hill, L.L. (2002). Feature type thesaurus. In *Alexandria Digital Library Project*. Available online at:
http://legacy.alexandria.ucsb.edu/gazetteer/FeatureTypes/FTT_metadata.htm.
- Hill, L. L. (2009). *Georeferencing: The geographic associations of information*. MIT Press.
- Hill, L. L., Frew, J., & Zheng, Q. (1999). Geographic names. *D-lib Magazine*, 5(1), 17.
- Janowicz, K. (2009). The role of place for the spatial referencing of heritage data. *The Cultural Heritage of Historic European Cities and Public Participatory GIS Workshop*, 9, 57.
- Janowicz, K., Schwarz, M., & Wilkes, M. (2009). Implementation and evaluation of a semantics-based user interface for web gazetteers. *Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW2009)*.
- Janowicz, K., & Keßler, C. (2008). The role of ontology in improving gazetteer interaction. *International Journal of Geographical Information Science*, 22(10), 1129-1157.
- Kazama, J., & Torisawa, K. (2007). Exploiting Wikipedia as external knowledge for named entity recognition. *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, 698-707.

- Keßler, C., Janowicz, K., & Bishr, M. (2009). An agenda for the next generation gazetteer: Geographic information contribution and retrieval. *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in Geographic Information Systems*, 91-100.
- Keßler, C., Maué, P., Heuer, J. T., & Bartoschek, T. (2009). Bottom-up gazetteers: Learning from the implicit semantics of geotags. *International Conference on GeoSpatial Semantics*, 83-102. Springer, Berlin, Heidelberg.
- Kim, J., Vasardani, M., & Winter, S. (2017). Similarity matching for integrating spatial information extracted from place descriptions. *International Journal of Geographical Information Science* 31(1), 56-80.
- Lambrix, P., & Tan, H. (2006). SAMBO—a system for aligning and merging biomedical ontologies. *Journal of Web Semantics*, 4(3), 196-206.
- Li, B., & Han, L. (2013). Distance weighted cosine similarity measure for text classification. In *International conference on intelligent data engineering and automated learning*, 611-618. Springer, Berlin, Heidelberg.
- Machado, I. M. R., de Alencar, R. O., de Oliveira Campos, R., & Davis, C. A. (2011). An ontological gazetteer and its application for place name disambiguation in text. *Journal of the Brazilian Computer Society*, 17(4), 267-279.
- Manguinhas, H., Martins, B., & Borbinha, J. (2008). A geo-temporal web gazetteer integrating data from multiple sources. *International conference on digital information management*, 146-153. IEEE.

- Mark, D. M., Smith, B., Egenhofer, M., & Hirtle, S. (2004). Ontological foundations for geographic information science. *Research Challenges in Geographic Information Science*, 335-350.
- Mark, D. M., & Turk, A. G. (2003). Landscape categories in Yindjibarndi: Ontology, environment, and language. In *International Conference on Spatial Information Theory*, 28-45. Springer, Berlin, Heidelberg.
- Martins, B., Manguinhas, H., & Borbinha, J. (2008). Extracting and exploring the geo-temporal semantics of textual resources. In *2008 IEEE International Conference on Semantic Computing*, 1-9. IEEE.
- McKenzie, G., Janowicz, K., Gao, S., & Gong, L. (2015). How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest. *Computers, Environment and Urban Systems*, 54, 336-346.
- Mostern, R., & Johnson, I. (2008). From named place to naming event: creating gazetteers for history. *International Journal of Geographical Information Science* 22(10), 1091-1108.
- Moura, T. H., & Davis Jr, C. A. (2014). Integration of linked data sources for gazetteer expansion. *Proceedings of the 8th Workshop on Geographic Information Retrieval*, 1-8.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT Press.
- O'Leary, D. E. (1998). Using AI in knowledge management: Knowledge bases and ontologies. *IEEE Intelligent Systems and Their Applications*, 13(3), 34-39.

- Popescu, A., Grefenstette, G., & Moëllic, P. A. (2008). Gazetiki: automatic creation of a geographical gazetteer. *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, 85-93.
- Purves, R., Edwardes, A., & Wood, J. (2011). Describing place through user generated content. *First Monday* 16(9).
- Raymond, C. M., Kytä, M., & Stedman, R. (2017). Sense of Place, Fast and Slow: The Potential Contributions of Affordance Theory to Sense of Place. *Frontiers in psychology*, 8, 1674.
- Regalia, B., Janowicz, K., Mai, G., Varanka, D., & Usery, E. L. (2018). GNIS-LD: Serving and Visualizing the Geographic Names Information System Gazetteer as Linked Data. *European Semantic Web Conference*, 528-540. Springer, Cham.
- Shvaiko, P., & Euzenat, J. (2005). A survey of schema-based matching approaches. *Journal on data semantics IV*, 146-171. Springer, Berlin, Heidelberg.
- Smart, P. D., Jones, C. B., & Twaroch, F. A. (2010). Multi-source toponym data integration and mediation for a meta-gazetteer service. *International Conference on Geographic Information Science*, 234-248. Springer, Berlin, Heidelberg.
- Smith, D. A. (2002). Detecting and browsing events in unstructured text. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, 73-80.
- Smith, B., & Mark, D. M. (2001). Geographical categories: an ontological investigation. *International Journal of Geographical Information Science*, 15(7), 591-612.
- Smith, B., & Mark, D. M. (2003). Do mountains exist? Towards an ontology of landforms. *Environment and Planning B: Planning and Design*, 30(3), 411-427.

- Southall, H., Mostern, R., & Berman, M. L. (2011). On historical gazetteers. *International Journal of Humanities and Arts Computing*, 5(2), 127-145.
- Sowa, J. F. (2000). Ontology, metadata, and semiotics. *International conference on conceptual structures*, 55-81. Springer, Berlin, Heidelberg.
- Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: principles and methods. *Data & knowledge engineering*, 25(1-2), 161-197.
- Sunna, W., & Cruz, I. F. (2007). Structure-based methods to enhance geospatial ontology alignment. *International Conference on GeoSpatial Semantics*, 82-97. Springer, Berlin, Heidelberg.
- Toral, A., & Munoz, R. (2006). A proposal to automatically build and maintain gazetteers for Named Entity Recognition by using Wikipedia. *Proceedings of the Workshop on NEW TEXT Wikis and blogs and other dynamic text sources*.
- Tuan, Y. F. (1977). *Space and place: The perspective of experience*. U of Minnesota Press.
- Van Assem, M., Menken, M. R., Schreiber, G., Wielemaker, J., & Wielinga, B. (2004). A method for converting thesauri to RDF/OWL. *In International Semantic Web Conference* 17-31. Springer, Berlin, Heidelberg.
- Vasardani, M., Winter, S., & Richter, K-F. (2013). Locating place names from place descriptions. *International Journal of Geographical Information Science*, 27:12, 2509-2532. DOI: 10.1080/13658816.2013.785550
- Wei, X., & Croft, W. B. (2006). LDA-based document models for ad-hoc retrieval. *In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* 178-185.

- Winter, S., Kuhn, W., & Krüger, A. (2009). Guest editorial: Does place have a place in geographic information science? *Spatial Cognition and Computation* 9(3), 171–173.
- Zhang, Z., & Iria, J. (2009). A novel approach to automatic gazetteer generation using wikipedia. *Proceedings of the 2009 Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources (People's Web)*, 1-9.
- Zhang, T., Shen, S., Cheng, C., Su, K., & Zhang, X. (2021). A topic model based framework for identifying the distribution of demand for relief supplies using social media data. *International Journal of Geographical Information Science*, 1-22.
- Zhong, C., Huang, X., Arisona, S. M., Schmitt, G., & Batty, M. (2014). Inferring building functions from a probabilistic model using public transportation data. *Computers, Environment and Urban Systems*, 48, 124-137.
- Zhu, R., Hu, Y., Janowicz, K., & McKenzie, G. (2016). Spatial signatures for geographic feature types: Examining gazetteer ontologies using spatial statistics. *Transactions in GIS* 20(3), 333-355.

APPENDICES

Appendix A: The Developed Ontology in OWL

The ontology file openly available in a public repository,

https://github.com/doh30/Geo_Ontologies/blob/main/gaz_onto_v1.0.owl

Appendix B: A total list of three factors (place affordances, events, key-descriptors) identified from Twitter data

Place Affordances	Key-descriptors	Events
Athletics	Apartment	Anniversary
Campaign	Bar	Basketball game
City planning	Burger place	College Football Playoff
Day activity	Coffee shop	Election
Drinking	Church	Football game
Driving	City	Gun violence
Eating	City hall	Homecoming
Entertainment	Clothing store	National Football League Draft
Game activity	Community	Rose Bowl
Healthcare	Creek	SEC Championship
Hiring	Dam	Softball game
News reporting	Department store	Sugar Bowl
Marketing	Downtown	
Morning activity	Furniture store	
Music activity	Garden	
Nightlife	Grocery store	
Party	Gym	
Physical activity	Landmark	
Political activity	Library	

Posting	Memorial hall	
Purchasing	Museum	
Recycling	Park	
Relaxing	Pharmacy	
Residential activity	Pool	
Running	Practice field	
Socializing	Residential area	
Sport activity	Restaurant	
Studying	River	
Taking photos	School	
Running business	Sheriff's office	
Training	Sports complex	
Traveling	Stadium	
Weather reporting	Store	
Weekday activity	Stream	
Weekend activity	Theatre	
	Waffle place	
	Weather station	