The Pennsylvania State University

The Graduate School

College of Earth and Mineral Sciences

A FRAMEWORK FOR DEVELOPING A GEOGRAPHIC ONTOLOGY FOR GAZETTEERS

A Thesis in

Geography

by

Kean Huat Soon

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2009

The thesis of Kean Huat Soon was reviewed and approved* by the following:

Douglas A. Miller
Associate Professor of Geography
Thesis Advisor

Todd S. Bacastow
Professor of Practice of Geography

Karl S. Zimmerer
Professor of Geography
Head of the Department of Geography

*Signatures are on file at the Graduate School

ABSTRACT

In applications such as web-based information retrieval, gazetteers are often used to provide information about places in the real world, including names, geographic coordinates, and feature class membership. Although this information is often sufficient to differentiate places with ambiguous geographic names for a given point in time, the information contained in standard gazetteers does not account for the dynamic aspect of geographic features that evolve over time. When utilizing gazetteers that describe geographic entities this shortcoming can cause two different places to be considered identical, or the same place to be viewed as different entities. Misidentifying places may cause significant mistakes in decision-making during crisis management or national security situations by missing critical information associated with a place of interest or by incorporating erroneous information.

To support the temporal continuity of geographical features, this thesis develops a geographic ontology that provides information regarding the evolution of coastal geographic features. This problem context is chosen because the changes in such features are strongly influenced by well-known earth system processes, such as sea level rise. Grounded in the notion of low-resolution conceptual neighborhoods, this thesis combines spatial data with the GeoNames gazetteer to develop the geographic ontology of coastal features. This ontology extends standard approaches by making explicit the knowledge of evolution of the spatial configuration of geographic features and their topological relationships. Using two local datasets from Florida and Alaska, the evaluation has shown that this type of geographic ontology can improve the automated matching between place names. This result in turn supports automated utilization of gazetteer-based information that represents the land surface at different times.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am very grateful to the following persons who have made this work possible: my wife Bee Er and daughter Philyra for their constant support, love and joy; my parents for praying for me all the time, my advisor Doug for his guidance and tolerance, my committee members, Frank and Todd for guiding me to think from a broader perspective.

CHAPTER 1 OVERVIEW

1.1 Motivation

Gazetteers have been widely used as a dictionary of geographic places to aid in information retrieval. A gazetteer includes the name, geographic coordinates and the feature class of a place (Hill 2006). Figure 1.1 is an example of place information that can be obtained from the Alexandria Digital Library (ADL) gazetteer (http://www.alexandria.ucsb.edu/) developed by the University of California at Santa Barbara.

Feature Name:

Mary Island – Jefferson County – New York - United States

Feature Class:

islands

Spatial Reference:

Geometry Type: Point

Long: -75.9217 Lat: 44.3653

Figure 1.1: An example of place information obtained from the Alexandria Digital Library.

This example shows that a place named "Mary Island" is located in Jefferson County, New York, United States. This place has point coordinates in latitude and longitude, and belongs to the island feature class. In general, one can not only obtain the geographic coordinates (which are often termed as geographic footprints, which are given either in point coordinates or a simple bounding box) of a place from a gazetteer, but one can also understand the nature of a place (e.g. an island) by referring to the appropriate class for that feature.

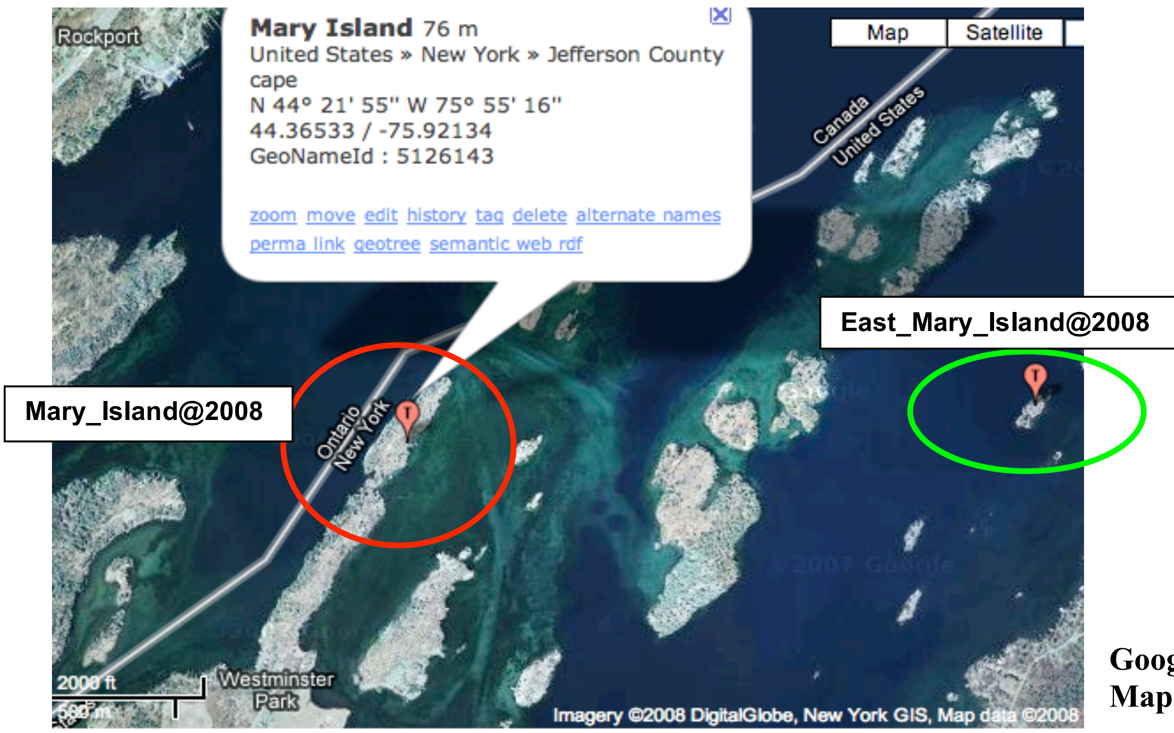
Place information from gazetteers like the one shown above is sufficient to disambiguate geographic place names that can appear in slightly different forms in various information sources, including maps and news stories (Smith and Crane 2001). The relationships between places in the current gazetteers are generally limited to generalization (e.g. Mary Island is an island) and aggregation (e.g. Jefferson County is part of New York). These types of relationships fall short of representing the dynamic aspect of geographic features, and the potential evolutionary changes of feature classes. Real world geographic phenomena are dynamic and often can change over time. When a given gazetteer may not cover all the places required in sufficient detail, utilizing multiple gazetteers becomes necessary (Hastings 2008; Janee 2006).

To illustrate problems that can arise when utilizing multiple gazetteers, Figure 1.2 depicts two maps representing the same area in Jefferson County, New York, United States. One is obtained from a recent Google Map via the GeoNames web service (http://www.geonames.org) and the other is a historical map dating back to 1908. By comparing these two maps, a few changes can be identified. The changes are as follows:

1) An island, which was originally named "Mary Island" (denoted as "Mary_Island @1908" on the historical map), has been renamed to "East Mary Island" (denoted as "East_Mary_Island @2008" on the Google map);

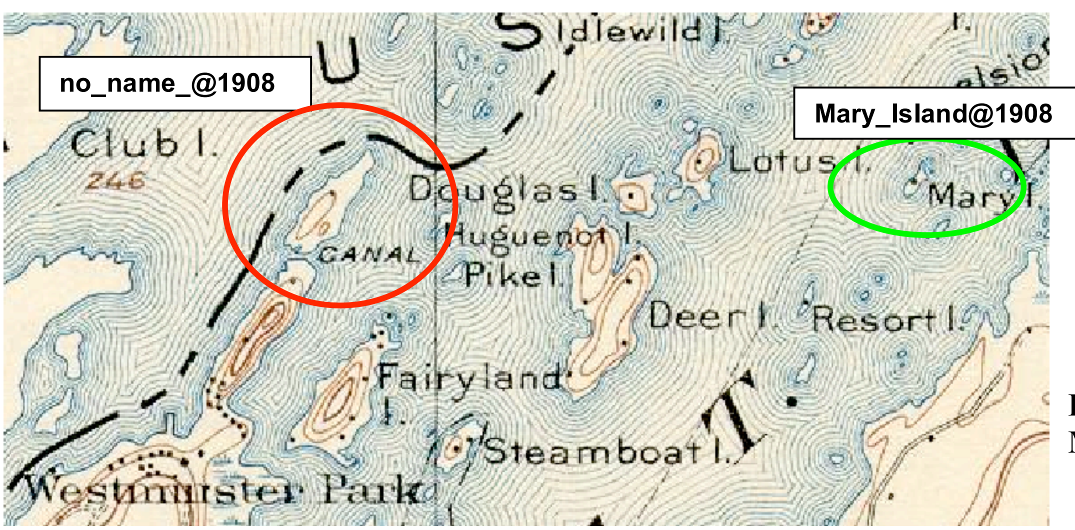
2) A place, which was originally unnamed (denoted as "no_name@1908" on the historical map), has been named "Mary Island" (denoted as "Mary_Island@2008") on the Google map;

3) "Mary_Island@2008", which was originally an island separated by a canal from the Westminster Park, has evolved into a cape.



Google Map

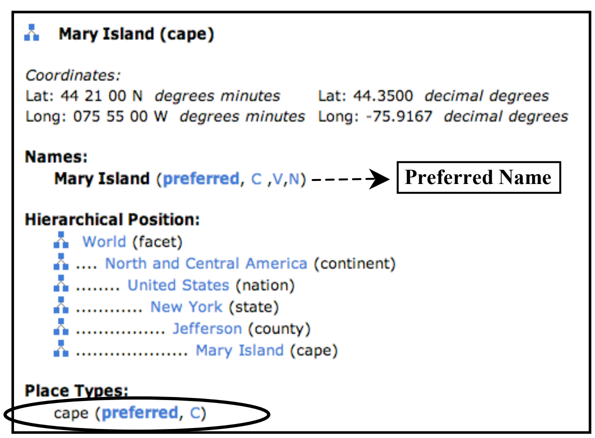
	Name	country	feature	km to center
1	Mary Island	United States	cape	1.07 km
2	East Mary Island	United States	island	1.37 km



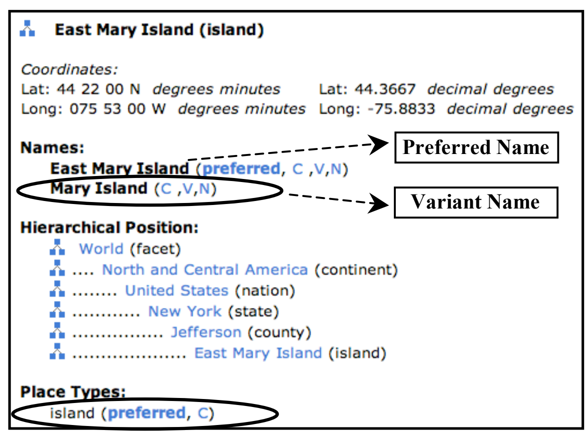
Historical Map

Figure 1.2: A Google map and a historical map illustrate the changes in naming and nature of "Mary Island" and "East Mary Island" in Jefferson County, New York, U.S..

In a current gazetteer, like the Getty Thesaurus of Geographic Names (TGN) (http://www.getty.edu/research/conducting_research/vocabularies/tgn/index.html) from the J. Paul Getty Trust, “Mary Island” and “East Mary Island” are described as follows. In Figure 1.3(a) “Mary Island” is referred to as a cape (circled), whereas in Figure 1.3(b) “East Mary Island” is referred to as an island (circled). Since the previous name of “East Mary Island” was “Mary Island”, the TGN also specifies “East Mary Island” as having a variant name called “Mary Island” (circled). A variant name simply means that the same place has more than one name. Variant name and preferred name are used either at different times or in different communities, but as its name implies, the preferred name is favored and it is often used. There can be more than one variant name, but only one preferred name is allowed for a place in gazetteers (J. Paul Getty Trust 2008). “Mary Island” for the cape, and “East Mary Island” for the island in the TGN are the preferred names. “Mary Island” for the island in the ADL gazetteer is another preferred name.



(a) Place information about “Mary Island”



(b) Place information about “East Mary Island”

Figure 1.3: Place information about “Mary Island” and “East Mary Island” from the Getty Thesaurus of Geographic Names.

Suppose that the TGN is to be utilized with the ADL gazetteer, the chance of mismatch[1] between the gazetteers can occur. The place information about “Mary Island” from the ADL gazetteer (as shown at the beginning of this section) can possibly be considered as the same with the

[1] A mismatch here means that two places that are essentially different are matched, but two places that are the same do not match. Matching wrong place information is critical as well as related information being not conflated results in information loss.

information from the TGN that describes “East Mary Island”, since the “East Mary Island” can also be referred to as “Mary Island”, and its feature class as island is exactly the same as the “Mary Island” from the ADL gazetteer.

Over time some places may have their names changed, and some others may be linked to different feature classes as a result of the evolution of geographic features. The reason that the “Mary Island” case above happened was that one gazetteer (i.e. the ADL gazetteer) still retains the original feature class, but the other gazetteer has changed the feature class to reflect the current state. When a gazetteer also includes the names that are previously used for a place, a mismatch like the “Mary Island” case occurs. Such a case can possibly happen to other coastal geographic features as well because these features are strongly influenced by earth surface processes such as sea level rise, erosion, and deposition over time.

A knowledge of geographic feature evolution is thus required to use multiple gazetteers. Computational utilization of multiple gazetteers with evolutionary knowledge can inform the information systems whether two features being compared are the same. When two places have the same preferred name and both are located within a certain proximity (e.g. located in the same county), even the places are referred to two different feature classes, the systems can still confidently treat the places as the same so long as one of the feature classes is an evolution of the other.

To supplement current gazetteers with such knowledge, the thesis develops a geographic ontology that takes into account the evolution of coastal geographic features as an example. Since there are many different types of natural and man-made places, this is a practical choice for constraining the problem domain for this initial implementation. The ontology is developed on top of the ADL Feature Type Thesaurus (FTT) (http://www.alexandria.ucsb.edu/gazetteer//FeatureTypes/FTT2HTM/FTT2ALPH.TXT). As its name implies, ADL FTT is a thesaurus that is used to classify places for the ADL gazetteer, and it is considered to be the richest classification system compared to other classification systems in current online gazetteers (Hill 2006). The other classification systems include those found in the Geographic Names Information System (GNIS) gazetteer (http://geonames.usgs.gov/domestic/feature_class.htm), the GeoNames web service (http://www.geonames.org/export/codes.html) and the TGN (http://www.getty.edu/research/conducting_research/vocabularies/guidelines/tgn_4_6_appendix_f_place_types.html). The knowledge that is currently represented in the ADL FTT, however, is limited to generalization and aggregation, which are not sufficient to deal with changes in geographic features over time (Janowicz and Keβler 2008).

With the assumption that people conceptualize geographic features topologically, the thesis utilizes topological relationships of coastal features defined within the ADL FTT as topological constraints (Egenhofer and Mark 1995). A place from the gazetteer therefore can be characterized with certain topological constraints depending on the feature class to which the place belongs. The ontology developed in this research is based on matching the gazetteer with spatial data. Provided the places from gazetteers can also have corresponding geographic representations (e.g. polygons) in spatial data, changes between the topological constraints of a place and the topological relationships of the corresponding geometry place may reveal a possible evolutionary relationship between geographic features. This can hold especially when the feature class of the place and the feature that the geometry represents are different (e.g. "lake" and "bay"). If the resulting topological changes conform to a formalization, then the feature that the polygon represents can be considered as a concept that has a reciprocal evolutionary relationship "evolves into" with the feature class, as another concept, in the geographic ontology.

To extend the formalization to include evolutionary topological constraints, the thesis considers the notion of Conceptual Neighborhoods from Egenhofer and Al-Taha (1992). Instead of taking the original eight topological relationships, the original version is reduced to four, which is referred to as "Low-Resolution Conceptual Neighborhoods." For feature classes like bay, when we come to the region where two water bodies (e.g. bay and ocean) connect, we often cannot tell to which water body the region supposedly belongs (Smith and Mark 1998). The topological relationships from the original Conceptual Neighborhoods that require a crisp boundary between two features (e.g. "meet") do not hold to represent such geographic features in gazetteers.

This leads to additional questions: how are the evolutions of geographic features formally specified? What are the fundamental characteristics of coastal geographic features? How to know if a geographic feature has evolved into a new feature? What are the possible features that a geographic feature can evolve into? What are the requirements that the new feature has to meet? In the following section, the objectives of the research are outlined to explore these questions.

1.2 Objectives

The research assumes that when a geographic feature evolves, its topological relationships with other surrounding features will consequently change as well. Through the formalization, one can use changes in topological relationships as indications of evolution of specific geographic features. Such indications in turn are used as a basis to develop the geographic ontology.

In this research, five key objectives are addressed:

1. Identify and characterize the fundamental characteristics of coastal geographic features

   A geographic feature can be characterized in different languages and cultures. Since all geographic features are bound in the geographic space, one can also differentiate a feature from other features with the topological relationships that the feature has with its surroundings. This objective aims to identify and characterize these inherent topological relationships (i.e. the topological constraints) for coastal geographic features. Since coastal geographic features usually are surrounded either partially (e.g. a bay) or completely (e.g. a lake) by another type of landform, vagueness exists on the “boundary” between two landforms of the same type (e.g. bay and ocean). This objective requires a consideration on vagueness to characterize the inherent topological constraints.

2. Formalize the evolutions of coastal geographic features

   This objective formalizes the evolutions of coastal geographic features. To allow for computational purposes in Geographic Information System (GIS), the existing methods for formalizing changes in geographic features have assumed that geographic features have crisp boundaries (Hornsby and Egenhofer 2000; Egenhofer and Al-Taha 1992; Burrough and Frank 1996). Building on one of these existing methods, the research seeks to formalize the evolutions of coastal geographic features, which have vague boundaries. In accordance with the topological constraints from the first objective, the resulting formalization should be based on topological change. Since topological changes can exist between two landforms of the same or different types, the formalization should also make clear between what types of landforms a given topological change creates.

3. Match footprints with spatial data to realize the possible changes in the real world

   When one matches a gazetteer with spatial data on the same place, the topological changes between the topological constraints of the place and the topological relationships of the respective geometry may indicate a possible evolution between geographic features. This objective explores this possibility by matching geographic footprints from a gazetteer with polygons in spatial data.

   To ensure the most accurate results, the spatial data should be "clean" in the sense that the positional errors of the data, such as "zigzag" and "overshoot", are already reduced to a minimum. One possible way is to use the spatial data from State GIS clearinghouses. State GIS clearinghouses provide spatial data for public use; positional errors in this data usually will be kept to minimum before releasing to the public.

4. Develop the geographic ontology

   This objective aims to develop the geographic ontology based on the results from the previous three objectives. To develop the ontology, as mentioned previously, if the topological changes conform to the formalization, then the feature that the geometry represents will be included in the ontology. To ensure that the conformation can be dealt with in an ad-hoc manner (i.e. automatic), this objective requires a computational implementation. In this implementation, the formalization should be represented as a set of rules, so that the conformation can be checked and the geographic ontology can be built automatically and represented in a computer-understandable language.

5. Evaluate the developed ontology

   An experiment should be conducted to evaluate the ontology developed from this research. This experiment is based on an information science perspective to test the effectiveness of the ontology in using multiple gazetteers computationally. As the FTT is the richest classification system currently available for gazetteers, the evaluation compares the ontology with the FTT in matching gazetteers.

1.3 Outline of the Thesis

This thesis is structured in such a way that the beginning chapters provide the background and conceptual understanding of this research. The middle chapters present implementation results based on the conceptual notions laid out in the preceding chapters. Lastly, the ending chapters evaluate the results found from the implementation, and conclude the findings. The following describe the contents of these chapters in more detail.

Chapter 2 contrasts some current work with the present research. It also offers some background contexts and delineates the notions that are used throughout this thesis. Particularly, this chapter discusses current research in ontology development, outlines the definition of geographic ontology and introduces the ADL Feature Type Thesaurus.

Chapter 3 defines the requirements for the spatial data to be used in this research, constructs the topological constraints, and develops the notion of low-resolution conceptual neighborhoods. This chapter sets forth the conceptual framework for the computational implementation in the following chapters. It bridges the conceptual and practical aspects of this thesis.

Chapter 4 computationally implements the conceptual framework outlined in the preceding chapters. It formally specifies the topological constraints in Web Ontology Language (OWL), creates the rules of conceptual neighborhoods, matches geographic footprints with spatial data, and develops the geographic ontology. More importantly, this chapter demonstrates two automatic systems that are developed in this research for developing the ontology and matching gazetteers.

Chapter 5 evaluates the developed ontology in comparison with the FTT. Place names from two local sources are considered for the evaluation. Place names from each of the local sources are attempted to match with place names from the GeoNames gazetteer using the FTT and the ontology. The results from the matching are compared and discussed.

Chapter 6 presents the findings, outlines the potential contributions from this research, and addresses the limitations. In particular, besides providing concluding remarks, this chapter takes a broader perspective on how to apply the developed ontology to assist in solving geographical problems, such as global warming. The chapter also discusses the limitations with providing recommendations for improvement.

Two peer-reviewed articles, as listed below, have been produced from Chapter 3. Other manuscripts focusing on the results from Chapter 4 and 5 are in preparation to submit to journals.

1) Soon, K. H. (accepted). Ontological Foundations for Geographical Data. In: Warf, B. (Ed.), The Encyclopedia of Geography. SAGE Publications, Inc.

2) Soon, K. H. (2008). Formalizing Geographic Features for Integrating Gazetteers. In: Cova, T.J., Miller, H.J., Beard, K., Frank A.U., and Goodchild, M.F. (Eds.), Conference Proceedings of GIScience 2008. 19-23 September 2008, Park City, Utah, U.S. (Poster)

CHAPTER 2 PREVIOUS WORK

2.1 Introduction

This chapter contrasts the research with the previous work in the geographic domain, and provides background information. In particular, Section 2.2 outlines recent research, Section 2.3 defines coastal geographic ontology, and Section 2.4 discusses the ADL Feature Type Thesaurus (ADL FTT).

2.2 Recent Research

There has been recent research on the development of ontology for gazetteers in the geographic domain (Goodchild and Hill 2008), but much of the unique complexities introduced by space and time have yet to be dealt with in a robust fashion. For example, ontologies as used by Purves *et al.* (2007) and Volz *et al.* (2007) are essentially reference gazetteers. When users define a query that contains a place name, their search engine will employ the geographic ontologies to provide a list of place names related to the place name in the query. After selecting one of these place names, the search engine will retrieve documents referring to the selected place name. Although this refinement resolves some ambiguous place names that appear in a query, only administrative entities are considered; places that belong to geographic features are left untouched. To be more comprehensive, their geographic ontologies should also include geographic features.

In his PhD dissertation entitled "A Geographical Ontology of Objects in the Visible Domain", Barry Bitters has built geographic ontologies that describe objects that humans can perceive in the real world (Bitters 2005). Although his ontologies are huge and cover a number of domains (e.g. transportation, agriculture, industrial, hydrography, recreation, earth's surface etc, http://vissim.uwf.edu/VOT_Ontology/Ontology.html), the relationships between concepts in the ontologies are limited. For example in the hydrography ontology, the concept "Bay" is only referred to as a "subClassOf" "BaysAndLagoons". This way of defining a concept is less comprehensive compared to the ADL FTT. In the ADL FTT, the concept "Bays" is defined as having "superClass" "hydrographic features", as being synonymous with "bahias", "coves" etc, and as being related to "bights" (http://www.alexandria.ucsb.edu/gazetteer//FeatureTypes/FTT2HTM/FTT2ALPH.TXT).

Although the ADL FTT is more comprehensive than Bitters' ontologies, the ADL FTT uses the single inheritance approach, meaning that a concept can only have one more general concept. This approach, as illustrated by Janowicz and Keβler (2008), falls short when dealing with concepts such as canal, which can be classified as a hydrographic feature as well as a manmade feature. Since the ADL FTT considers "hydrographic feature" and "manmade feature" as two separate major classes, a concept that is a sub-concept of hydrographic feature cannot be at the same time a sub-concept of a manmade feature. Although Janowicz and Keβler (2008) have developed a geographic ontology to overcome this shortcoming, the knowledge of how two concepts may evolve is not considered in their ontology. In other words, they regard only generalization and aggregation relationships between concepts. In their example, a canal *is* a manmade feature and hydrographic feature, and a canal also *has* a destination (i.e. it connects with another hydrographic feature), they do not take into account whether a geographic feature can possibly evolve into another feature, for example could a lake evolve into a stream?

Geographic ontologies have also been actively developed in government agencies such as DAML (DARPA Agent Markup Language, http://www.daml.org/) and the Ordnance Survey of Great Britain (http://www.ordnancesurvey.co.uk/). Taking the Ordnance Survey as an example, this government agency aims to mitigate the issue of semantic heterogeneities between the understandings of their customers and spatial data (Mizen *et al.* 2005). To achieve this aim, the Ordnance Survey has developed geographic ontologies (which can be downloaded from http://www.ordnancesurvey.co.uk/oswebsite/ontology/) to explicitly define the semantics of the spatial data. These ontologies are generally defined in five relationship types - taxonomic, synonym, topological, mereological, and affordance. The instances of these types are: for taxonomic – "is a"; for synonym – "same as", "local to", "language"; for topological – "next to", "empties", "feeds from", "bounds", "bounded by", "contained in", "contains", "built across"; for mereological – "part of", "has part"; and finally for affordance – "affords", "may afford" (Hart *et al.* 2004). From these relationship types, obviously none of them can be used to describe the relationship between two geographic features that can evolve. Coastal features, for example, evolve all the time.

There has been substantial work on modeling coastal evolution in coastal geomorphology (Carter and Woodroffe 1994; Lakhan 2004). Most of this work, however, relies on quantitative analysis, and requires expert knowledge to develop models like the Markovian Inheritance (Leont'yev 2003; Cowell and Thom 1994). Markovian Inheritance consists of a tree, called a Markov tree, in which each node represents a state or essentially a coastal feature like a barrier or a mainland beach. Stochastic environmental inputs, such as sediment transports and bed elevation, are required to transit from one state to another. In other words, two coastal features would not be related if the required inputs do not achieve certain thresholds set by the experts. In contrast, the present research develops a framework based on human conceptualization or "common-sense knowledge". Such a conceptualization does not necessarily require lengthy experience or formal education. It is developed from the interaction between people and the environment (Egenhofer and Mark 1995).

In the GIScience domain, Raper and Livingstone (1995) have developed a model called OOgeomorph, which enables specialists (e.g. geomorphologists) to model geomorphologic processes (e.g. deposition, erosion), coastal features (e.g. spits), and attributes (e.g. tide level). Their research combines objects, processes and spatial data into one single model, where in previous studies process modeling and object modeling were often treated separately. Their model, however, depends on expert knowledge from the specialists.

Molenaar and Cheng (2000) have developed a formalism to reason about the dynamics of coastal features. The formalism is based on the displacement and overlap of two corresponding regions from images to determine whether a coastal feature has "shifted", "split", "merged", "expanded", "shrunk", "appeared", and "disappeared" into another feature. As an example, by matching two images from different times on the same place, if the area of the location from an older image is smaller than the area of the same place from the other newer image, the formalism will determine that the feature that the region represents in the older image has expanded into another feature. Their approach, however, seems to depend on the method used to produce the images, i.e. different classification systems (e.g. supervised or unsupervised) may result in different sizes and locations of the same feature for the same image. Thus a feature that is determined as having expanded by a given method could possibly be determined smaller by another different method.

The idea of grounding geographic terms in spatial data is not novel. Third *et al.* (2007) have grounded geographic terms, such as rivers and lakes, in spatial data by looking at the shape and measurement of a polygon. For example, a polygon that represents a river will usually have a width that is much smaller than its length, compared to a polygon that represents a lake. In this way, Third *et al.* (2007) claim that geographic terms can be defined more precisely in order to integrate spatial data. In supplementing their work, the present research asserts that by looking at the topological relationships of a polygon, one can also describe closely the characteristics of the feature that the polygon represents. For example, a polygon representing a lake will usually be disjoint from other polygons representing water bodies. A "river" polygon will usually meet with other "water body" polygons. In specifics, the present research is limited to the topological relationships of a polygon, which corresponds to a particular place from the gazetteer. If the topological relationships differ from the topological constraints of the place, the changes may signify the evolution between the feature class of the place and the feature that the polygon represents in the real world.

2.3 Coastal Geographic Ontology

This research focuses on *coastal geographic* ontology. Fonseca *et al.* (2006) specify a geographic ontology as an ontology that describes geographic entities that are "either constructed features or natural variations on the surface of the Earth" (p. 314). The geographic ontology in this research only concentrates on macro-coastal geographic features, such as bays or lakes, which have proper names in the real world and are subject to coastal geomorphologic processes. The ontology hence does not consider constructed features such as buildings or inland features such as mountains.

One can also regard the ontology in this research as belonging to the field of "Ethnophysiography", the term coined by David Mark and Andrew Turk for the meaning of a science of topography (Mark and Turk 2003). But instead of looking at different cultures and languages that can influence the definitions of geographic features, as Mark and Turk (2003) have developed for the distinctions between Australian and Yindjibarndi, this research focuses only on English terms. In summary, the focus of this research is not to investigate the semantic differences between geographic feature terms owing to different languages and cultures, but instead to investigate the spatial relationships between geographic features that make the terms different. For example, "lake" and "bay" both can be referred to as hydrographic features, but because these two features have distinct spatial relationships with other features in the real world, such relationships make the two terms different.

2.4 Alexandria Digital Library Feature Type Thesaurus (ADL FTT)

According to the ADL FTT, all features that exist in the real world and that are named can be classified into six major classes. These classes are *administrative areas*, *hydrographic features*, *land parcels*, *manmade features*, *physiographic features*, and *regions*. A list of feature classes for the ADL FTT can be found here http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/top.htm.

Generally, the ADL FTT includes four relationships to relate feature classes. As an example, Figure 2.1 describes a small portion of the thesaurus focusing on hydrographic features. These four relationships are “Broader Term (BT)”, “Narrower Term (NT)”, “Used For (UF)”, and “Related Term (RT)”. The “Broader Term” and “Narrower Term” relationships can be considered similar to the inheritance (generalization) and composition (aggregation) relationships in the object-oriented sense (e.g. “bays” –“Broader Term” -> “hydrographic features”; “bays” –“Narrower Term” - > “fjords”), while the “Use For” and “Related Term” relationships respectively refer to synonym and “something being related” to a particular term. Compared to the other three relationships, “Related Term” is defined loosely. If two terms are related according to one’s view, one can designate them as a “Related Term”.

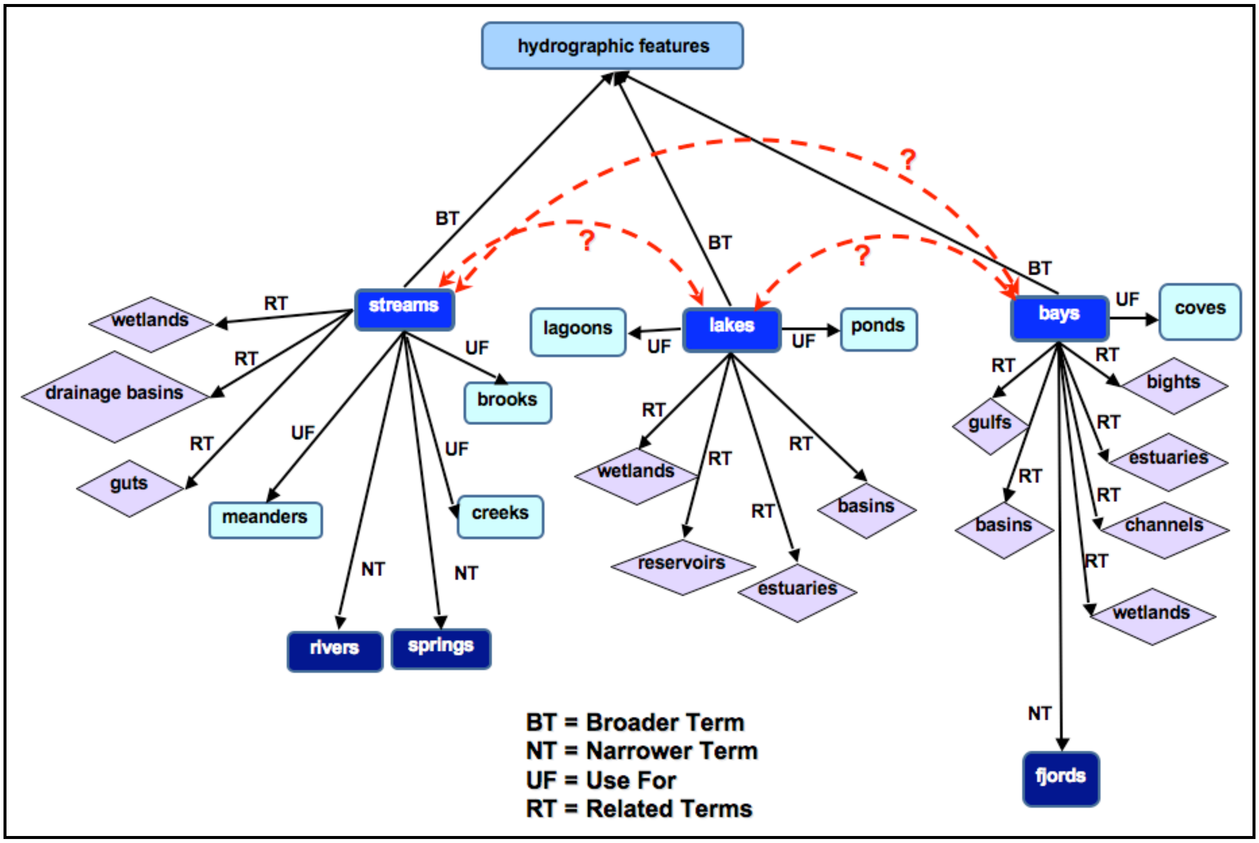


Figure 2.1: A portion of the ADL FTT that focuses on "streams", "lakes", and "bays".

Instead of taking all the feature classes from the ADL FTT, this research only considers hydrographic features such as bays, lakes, and streams. The reason for choosing hydrographic features is twofold. First, hydrographic features topologically are similar to their land feature counterparts. For example islands and lakes are usually located *inside* the landform of a different type. Bays and capes both always *connect* with the landform of the same type. The consideration of hydrographic features only should suffice to develop an ontology conforming to the formalization. Second, the research uses spatial data from public domains, such as state spatial data clearinghouses. Such data sources only describe hydrographic features in polygons; no polygons are available for land features such as capes and deltas. Spatial data in polygons are important for discerning topological change between geographic features.

In Figure 2.1, note that "streams", "lakes", and "bays" are preferred terms according to the ADL FTT. The preferred terms here mean that these terms are preferably to be used when referring to similar terms. For example, when referring to a lagoon or a pond, the term "lakes" is used for classifying places in the ADL gazetteer. Figure 2.1 only illustrates some portions of the terms that are related to the three feature classes. The complete ADT FTT can be obtained from http://www.alexandria.ucsb.edu/gazetteer//FeatureTypes/FTT2HTM/FTT2ALPH.TXT. As described in the metadata of the thesaurus (http://www.alexandria.ucsb.edu/gazetteer/ FeatureTypes/FTT_metadata.htm), this complete version contains 210 preferred terms and 1046 non-preferred terms in total.

Although the example illustrated in Figure 2.1 is straightforward, it demonstrates one limitation of the ADL FTT. As one can see, no direct relation is created among "streams", "lakes", and "bays" (the red dashed arrows indicate the missing links). These features are only related through the links with "hydrographic features", because all of them are hydrographic features. This simplistic way of representing the relationships between features assumes that the features are conceptually equivalent under one major class. Such a representation would raise an issue when the representation is based on integrating two places that have the same name and are located in the same county and state, but one of them is referred to as a waterfall, which is identified as a hydrographic feature in the ADL FTT, and the other is a lake. In such a case, the two places could possibly be matched, as both are hydrographic features.

Within the same major class a feature sometimes can be more related to another feature, for instance from an evolutionary aspect. Consider the relationships between a lake and a bay. A lake could possibly evolve into a bay if the sea level has risen. A lake, under most circumstances, never evolves into a waterfall. Therefore, the two places as mentioned in the example above should not be matched because a lake is not likely to become a waterfall. The two places may be close to each other spatially but they are not the same and one will not likely evolve into the other.

A detailed treatment is needed to define the relationships among features within a major class. As the main focus, this research project formulates the evolutionary relationship between hydrographic features based on topological change. The objective is to extend the thesaurus to utilize multiple gazetteers as the geographic features evolve over time.

CHAPTER 3 DESIGN AND IMPLEMENTATION

3.1 Introduction

This chapter discusses the design and implementation of the framework. The next chapter will present the results from the implementation outlined here. Section 3.2 outlines the requirements for the spatial data used to develop the ontology. Section 3.3 presents the conceptual design of the framework and demonstrates some general steps to computationally implement the framework.

3.2 Spatial Data Requirements

Spatial data plays a vital role in providing concepts for the ontology. To ensure that the resulting ontology describes concepts that are necessary, yet comprehensive, the spatial data to be considered in this research should meet certain requirements. These requirements are described as follows:

1) The ontology is developed based on topological change. To realize the topological changes between geographic features, the spatial data has to be represented in polygons with each polygon representing a particular geographic feature in the real world. This thesis uses *feature name* to refer to the feature a polygon represents in the spatial data, while the term *feature class* is reserved for a place from the gazetteer.

2) The geographic features that the polygons represent should be as complete as possible in order to create a more comprehensive ontology – i.e. the more the geographic features are represented by the spatial data, the more the concepts will be included in the ontology.

3) To reduce positional errors that may exist in the spatial data, the data should be obtained from a reliable source and should be checked use.

3.3 Research Plan

The plan of this research can be divided into two major phases – the design phase and the implementation phase. The goal of the design phase is to design the framework, particularly in constructing the topological constraints and the notion of low-resolution conceptual neighborhoods. The implementation phase is to computationally construct and test the ontology. The following subsections describe each of these phases and their sub-phases in greater detail.

3.3.1 Design Phase

This section begins with the formulation of the topological constraints for geographic features, followed by discussion of the development of low-resolution conceptual neighborhoods. Initial test results are presented to demonstrate how the formalization can be applied to the GeoNames gazetteer and the spatial data.

3.3.1.1 Topological Constraints

As people tend to organize geographic space topologically (Egenhofer and Mark 1995), this thesis assumes geographic features inherently have certain topological relationships related to human conceptualizations. For instance, when we think of a lake, we often understand it as a feature that is located "inside" of another land feature. Likewise, when we talk about a bay, we often conceptualize it as a feature that "overlaps" with another water body. We usually do not think of it as being "disjoint" with, "inside" of, or "meet" with the water body. Especially for "meet", since no crisp boundary exits between a bay and the other water body, we often cannot tell how a place should be referred to when located at the region between the bay and the water body (such a region is highlighted in gray in Figure 3.1).

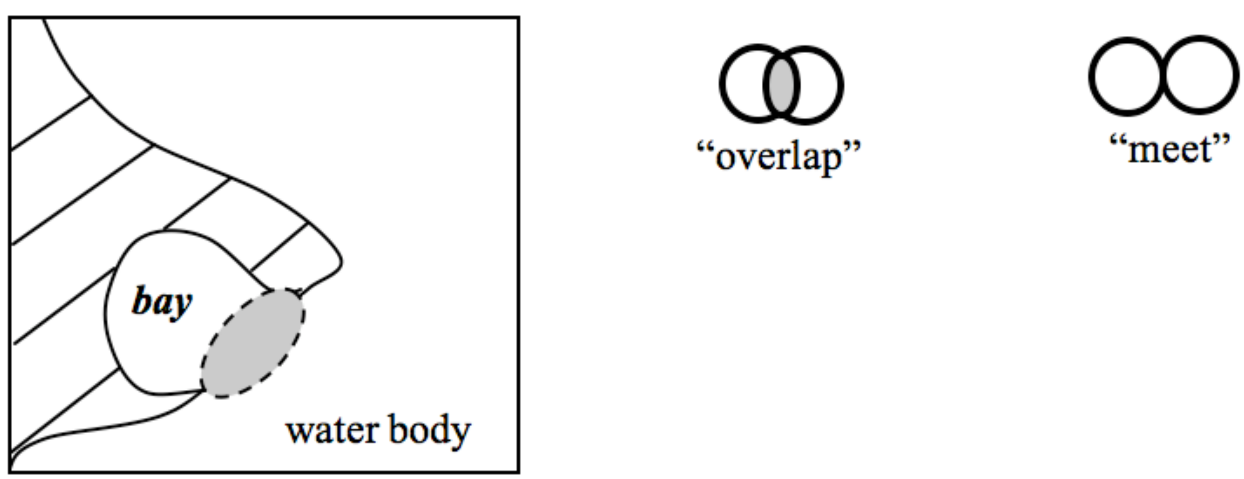


Figure 3.1: A bay "overlaps" with the water body.

Based on the characterizations above we can generally classify natural hydrographic features into two main categories. The first category includes hydrographic features that have "inside" as their topological constraint. Examples of hydrographic features that belong to this category are pond, reservoir, and lake. The second category refers to hydrographic features that have "overlap" as their topological constraint. This category includes bay, gulf, and stream.

3.3.1.2 Low-Resolution Conceptual Neighborhoods

Since changes in coastal features are continuous, the evolution of coastal features can be related to changes in topological relationships. To reason about topological changes between geographic features, the notion of conceptual neighborhoods (Egenhofer and Al-Taha 1992) has been adapted from the spatial reasoning literature. Cohn and Hazarika (2001) provide a review in spatial reasoning. In conceptual neighborhoods, as illustrated in Figure 3.2, two adjacent topological relationships are considered as *neighbors*. For example, "disjoint" and "meet" are neighbors, "meet" and "overlap" are neighbors, but "disjoint" and "overlap" are not neighbors. Given this, according to conceptual neighborhoods, two features that are previously "disjoint" can only "meet" before they can "overlap".

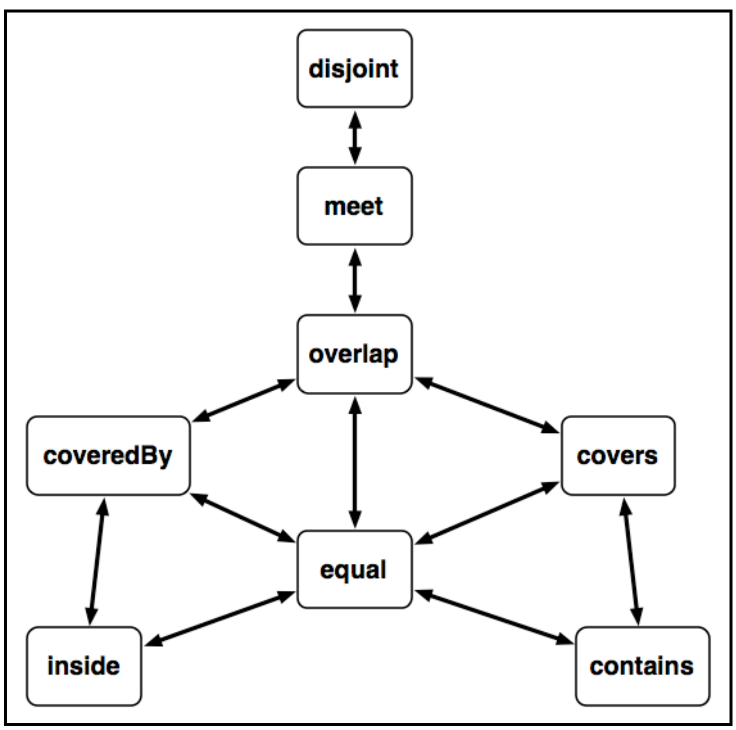


Figure 3.2: The original conceptual neighborhoods from Egenhofer and Al-Taha (1992).

The original conceptual neighborhoods however are more suitable for representing topological changes between spatial geometries (e.g. polygons), which have a precise boundary. This notion of conceptual neighborhoods may not closely apply to geographic terms that are used as feature classes in gazetteers. In this research, a modified version of the conceptual neighborhoods called "Low-Resolution Conceptual Neighborhoods" is developed (see Figure 3.3).

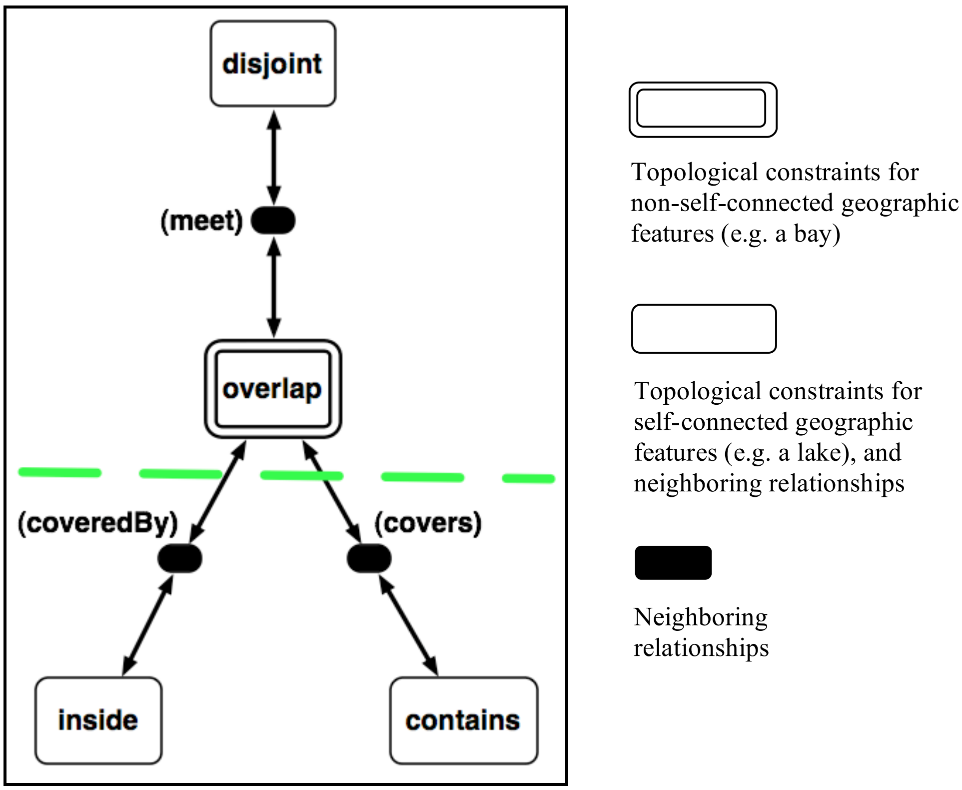


Figure 3.3: Low-resolution conceptual neighborhoods.

The low-resolution conceptual neighborhoods are developed based on the topological constraints outlined in the previous section. As one can see from Figure 3.3, the number of topological relationships is reduced to four from the original eight. The remaining topological relationships are "disjoint", "overlap", "inside", and "contains". Of these relationships, "disjoint", "inside" and "contains" represent the topological constraints for geographic features that are self-connected (e.g. lake, reservoir). All self-connected features belong to the first category mentioned in the previous section.

Note that the topological relationship "contains" here is symmetrical with "inside" instead of giving the meaning that a specific feature contains another smaller feature. This research assumes that humans conceptualize a self-connected feature as a simple region without holes. A lake could also contain an island, but in general we understand a lake as simply a water body that contains water. Therefore, in most cases, this thesis assumes "contains" as a topological constraint without stating it explicitly.

The "equal" relationship is also omitted here. According to Egenhofer and Mark (1995), people tend to characterize geographic space as two-dimensional space, so no geographic feature (e.g. a bay) can be characterized as *exactly* overlapping another geographic feature. Different from manipulable table-top objects such as tables and boxes, a box can exactly overlap a table by putting the box, that is large enough to exactly occupy the whole table top, directly on the table surface.

The topological relationship in the double-lined rectangle (i.e. "overlap") refers to geographic features that are non-self-connected (e.g. bay, gulf) (i.e. the second category). For these features, we usually conceptualize them as being connected with another similar landform (e.g. a bay usually connects with the sea).

The low-resolution conceptual neighborhoods can also be divided into two groups, which are delineated with a green-dashed line in Figure 3.3. The upper group includes the topological relationships between geographic features that belong to the same type, for example, a bay and the sea are both water features, while the lower group describes the topological relationships for landforms with different types (e.g. a lake and the surrounding land (water and land features)).

Diverging from the original version, "disjoint" and "overlap" (or "inside" and "overlap", or "contains" and "overlap") are neighbors in this modified version. These neighborhoods allow for representing the evolution in coastal geographic features that do not require identification of the crisp boundary between coastal features. Following this, a self-connected feature can evolve into a non-self-connected feature by changing its topological relationship from "disjoint" (or "inside") to "overlap", and vice versa (e.g. "lake" and "bay").

Besides the four topological constraints, in Figure 3.3 one can see the other three small-filled-round rectangles. The relationships represented in these 3 small round rectangles are the neighboring relationships. In order to ensure whether a topological change signifies an evolution between geographic features, the topological relationships of the polygon should meet the neighboring relationships, and the feature name of the polygon should be different from the feature class.

To be specific, the topological constraints of a place determine the required neighboring relationships that a corresponding polygon should meet. For example, if the feature class of a place is lake (or other self-connected features), which has topological constraints – "disjoint" and "inside", the required neighboring relationships that the corresponding polygon should meet are "meet" and "coveredBy" only. In other words, the polygon should touch other polygons representing water bodies (because "meet" and "coveredBy" signify contact with other polygons). If the feature class is bay, a non-self-connected feature, then the required neighboring relationships are "meet" and "coveredBy", *or* "disjoint" and "inside". In this case, the polygon can either touch or not touch other polygons representing water bodies.

3.3.1.3 Initial Testing

Two examples are now provided to show matching footprints to polygons in a real dataset based on the notions of topological constraints and low-resolution conceptual neighborhoods discussed above. The purpose here is to give the reader working examples of how the formalization can be applied to develop the ontology. Please keep in mind that the matching and the creation of the ontology in this research are machine-generated using the ontology development system.

Figure 3.4 illustrates a matching for a place named "Point Judith Pond", which is located in Washington County, Rhode Island. In the map on the left, the green dot is the geographic footprint of the place in point coordinates from the GeoNames gazetteer, and the blue complex polygon that the footprint falls in represents the corresponding geometry from the spatial data. As this place is referred to as a lake in the gazetteer, the place has topological constraints "disjoint" and "inside".

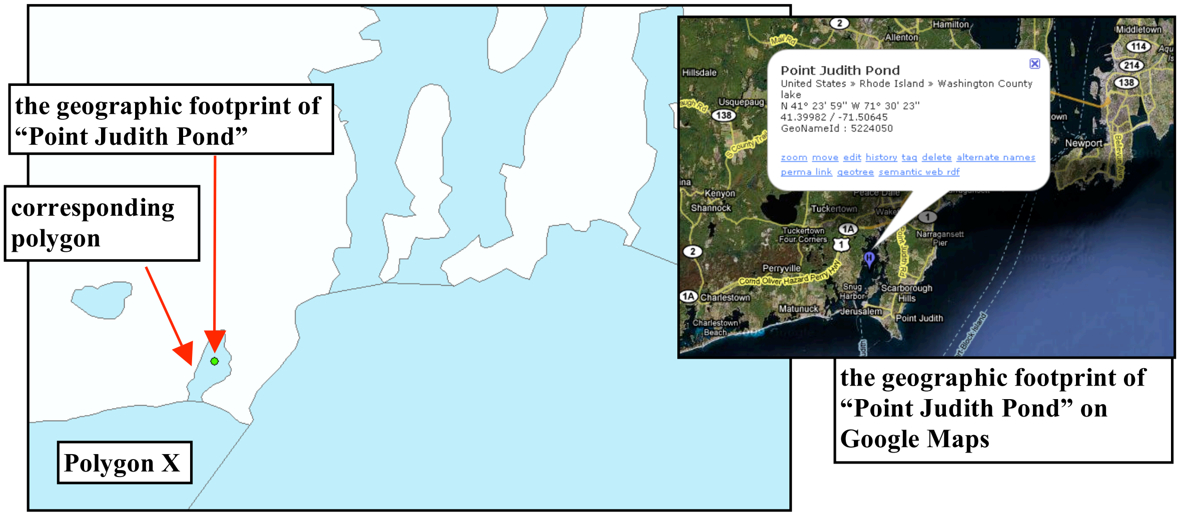


Figure 3.4: A match between a geographic footprint from the GeoNames gazetteer and a corresponding spatial geometry.

On the right of Figure 3.4, a screenshot, which is obtained from the GeoNames website, shows the same footprint matched on a Google map. As the GeoNames web service has been incorporated with Google™ Maps, each footprint found from the web service can be plotted on a Google map. In this research, this facility is used to verify the matching between footprints and polygons. For example, from the Google map on the right, one can verify that the matching between the footprint and the polygon on the left is correct.

From the spatial data, the feature name of the polygon is obtained. Knowing that the feature name is “bay”, which is different from the feature class, the topological relationships of the polygon are checked to verify the required neighboring relationships. As shown in Figure 3.4, the complex polygon touches with another “blue” polygon (denoted as “Polygon X”), which is also a water body, “bay” (the feature name) is created and related with “lake” (the feature class) via the evolutionary relationship “evolve_into” in the ontology.

The example above is just one of many potential situations. For example, a lake can also evolve into other geographic features, such as a stream. Figure 3.5 demonstrates a second example when place information from the GeoNames gazetteer is matched with spatial data. According to the GeoNames gazetteer, a place named “Cow Island Pond”, which is located in Norfolk County, Massachusetts is a lake. When the geographic footprint of “Cow Island Pond”, indicated as a green point in Figure 3.5, is matched with the spatial data the polygon that the footprint falls in touches other water body polygons. Since “meet” is the required neighboring relationship in this case, and the polygon represents a stream, the concept “stream” is added to the ontology. As a result, by combining the two examples above, three concepts are created in the ontology with “evolves_into” reciprocal relationships connecting the concepts (see Figure 3.6).

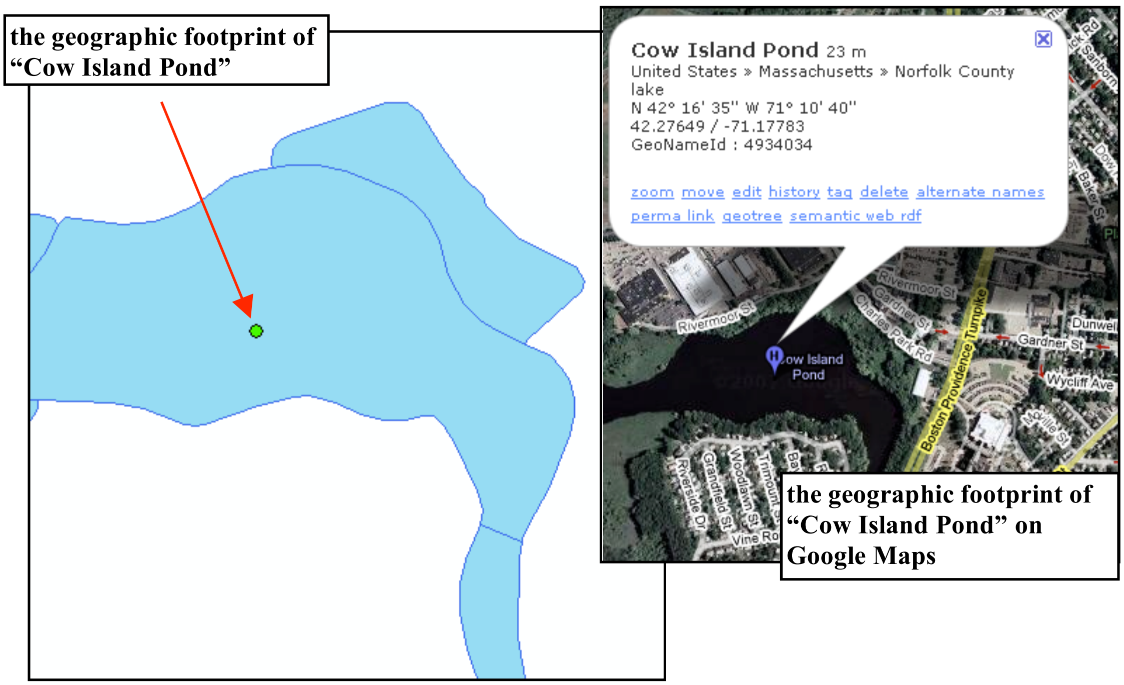


Figure 3.5: A second example of matching a geographic footprint with a corresponding spatial polygon.

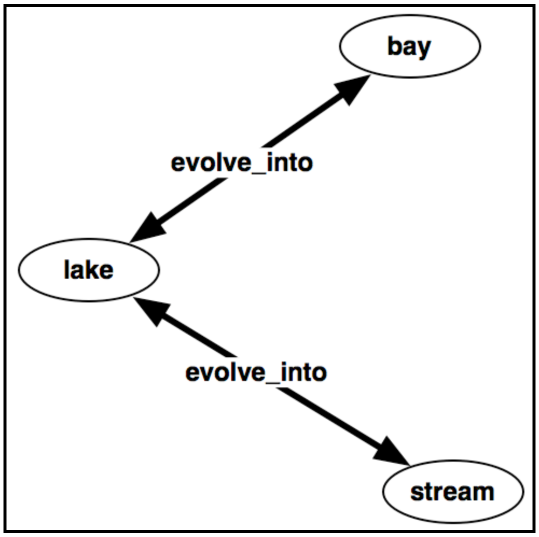


Figure 3.6: A simple geographic ontology that describes the evolutions between coastal features.

3.3.2 Implementation Phase

This phase aims to computationally implement the framework and then test the geographic ontology. The following steps are required:

1. Creating an initial ontology with fundamental concepts supported by topological constraints and engineering the rules of low-resolution conceptual neighborhoods;
2. Retrieving place information from the gazetteer;
3. Matching geographic footprints with polygons in the spatial data, and retrieving topological relationships between polygons;
4. Developing the geographic ontology in the Web Ontology language (OWL);
5. Evaluating the developed ontology.

Figure 3.7 illustrates the implementation phase workflow. The numbers in the red circles show the order of the important steps in this phase. To begin with the first step, an initial ontology is created with concepts associated with the topological constraints. This initial ontology is necessary at this very beginning stage because it allows the system to check the topological constraints of a feature class when the matching with the spatial data is carried out.

After the initial ontology is created, the rules of low-resolution conceptual neighborhoods are engineered. With the topological constraints and the rules in place, the matching with the spatial data can then be implemented. To perform the matching, place information from the GeoNames gazetteer is retrieved and matched with the hydrography layer from the MassGIS state clearinghouse.

If the geographic footprint falls in one of the polygons, the feature name and the topological relationships of the matched polygon are checked with the rules of low-resolution conceptual neighborhoods and the topological constraints. If the polygon meets the requirements, the feature name will be added to the initial ontology with an evolutionary relationship.

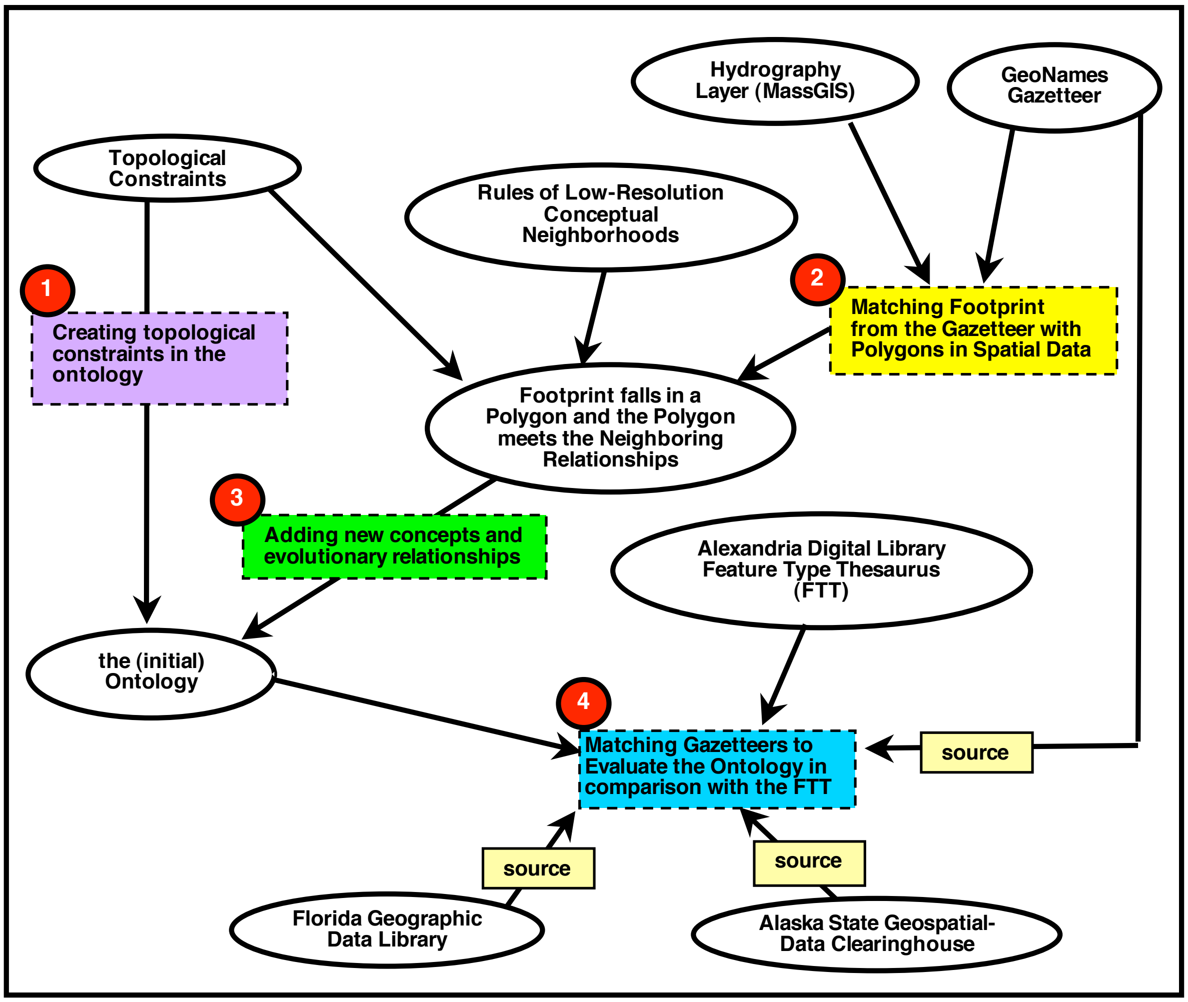


Figure 3.7: The workflow of the implementation phase.

Ultimately for the evaluation, the ontology is compared with the FTT for matching place names from three sources, namely GeoNames, the Alaska State Geospatial-Data Clearinghouse (ASGDC) and the Florida Geographic Data Library (FGDL). Because Alaska and Florida states in the U.S. have the longest and second longest coastlines, place names from these two states are considered. Place names from ASGDC and FGDL in turn were originally obtained from the gazetteers Alaska Place Name Dictionary and USGS Geographic Names Information System, respectively. As the place names from these two sources cover different regions, a global gazetteer, which can be used to match with each of these sources, is needed. GeoNames as a global gazetteer is chosen.

In the evaluation, a list of place names, which also includes the names of the county and the state the place is located in, is checked with two sources at a time. For each place, if the same names appear but the place refers to different feature classes in both sources, the place will be considered for matching, and the two different feature classes will be treated as one mismatched category. The result from the evaluation is determined between the numbers of mismatched categories that can be individually matched by the ontology and the FTT. The following sections explain each of the steps involved in more detail.

3.3.2.1 Creating the Topological Constraints in an Initial Ontology and Engineering the Low-Resolution Conceptual Neighborhoods Rules

To apply the notions developed in the design phase, the topological constraints are defined in an initial ontology for three fundamental concepts - *stream*, *bay* and *lake*. These three concepts are chosen because they represent three fundamental differences for hydrographic features: a stream usually connects to two (or more) water bodies, a bay usually connects to only one water body and it is partially enclosed, and a lake usually is located completely inside of a land feature. The schematic diagram in Figure 3.8 depicts the fundamental differences.

Besides the topological constraints, in the initial ontology the three fundamental concepts are described as specializations or subclasses of a “hydrographic feature”. Also, as the concept “lake” is topologically related to land feature, two additional concepts – “land region” and “physiographic feature” are also defined in the initial ontology. Instead of only creating one concept “land feature”, which can already meet the need, the initial ontology defines two concepts - “land region” and “physiographic feature”. This ensures that the final ontology, which is developed from this initial ontology, is only different from the FTT by the evolutionary relationships. The ontology does not contain more or different concepts than those found in the FTT.

The ontology is represented in Web Ontology Language (OWL) (http://www.w3.org/TR/owl-features/), the most favored machine language for ontology. Representing the ontology in OWL allows for reasoning among feature classes using the existing reasoning engines such as Jena (http://jena.sourceforge.net/), which is an Open Source Java framework for semantic web applications.

Figure 3.9 describes the pseudo-code for the rules of low-resolution conceptual neighborhoods. This pseudo-code is implemented in the Java Programming Language and is linked with Jena to obtain the topological constraints of feature class from the initial ontology. Together with the topological constraints and the rules of conceptual neighborhoods, the system developed for constructing the ontology can automatically determine whether the feature name of the polygon that the footprint falls in should be included in the ontology.

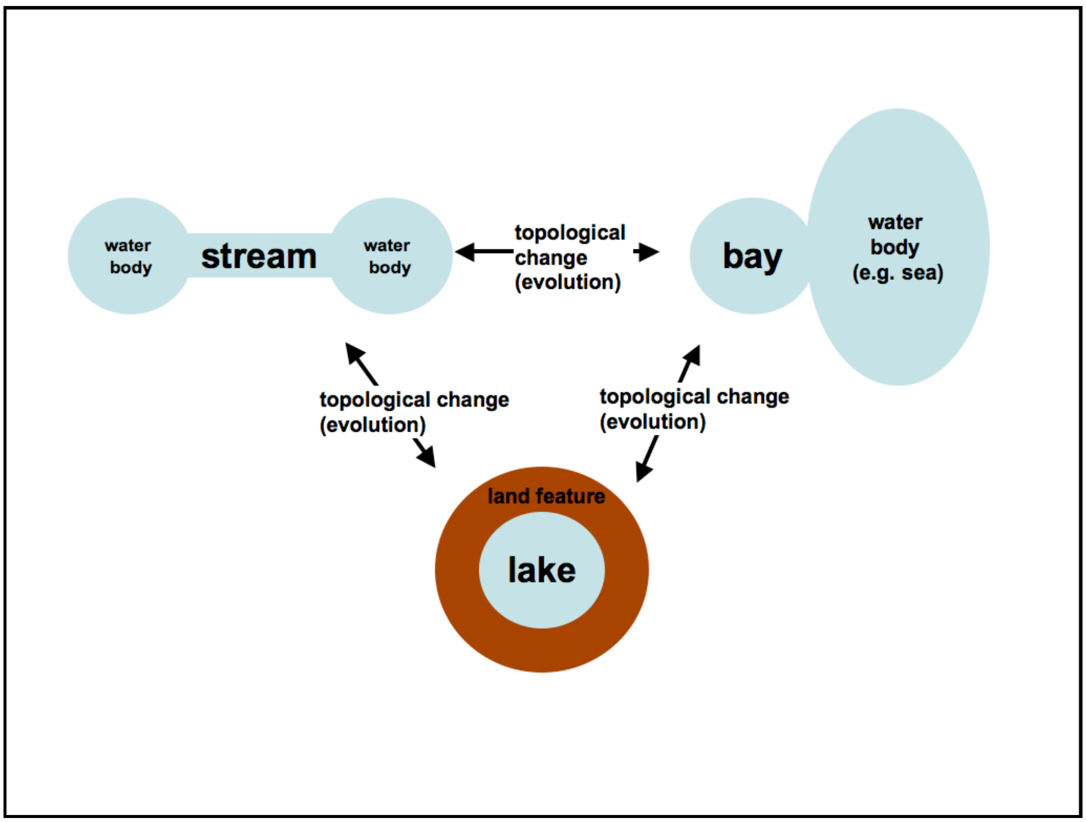


Figure 3.8: Three fundamental concepts – *stream*, *bay*, and *lake*. The evolution from one concept to the other signifies topological change.

```
if the footprint falls in a polygon & feature name different from
   feature class

    get number of touches for the polygon
    get topological constraint of the feature class

    if number of touches >0 & topological constraint = "inside" then
        include feature name
    else if number of touches >0 & topological constraint = "overlaps" then
        include feature name
    else if number of touches =0 & topological constraint = "overlaps" then
        include feature name
    else
        do nothing

else
    do nothing
```

Figure 3.9: The pseudo-code for the rules of low-resolution conceptual neighborhoods.

3.3.2.2 Retrieving Place Information Using the GeoNames Java Client

GeoNames is chosen in this research because it provides a web service where the end users can easily retrieve place information. Another option would be the ADL gazetteer, but unfortunately at the time of writing, the gazetteer is facing technical difficulties as the institution that manages the gazetteer is in the process of resurrecting it. The Getty Thesaurus of Geographic Names (TGN) has a web service, but their web service is only available for subscribers.

To retrieve information about places from the GeoNames gazetteer, the Java Client publicly available from GeoNames (http://www.geonames.org/source-code/) is used. As its name implies, the Java Client is written in the Java Programming Language. To obtain place information, one can write a computer program to make use of the Java Client without the need to separately parse the resulting output from the web service. As opposed to USGS Geographic Names Information System (GNIS), this gazetteer does provide a web service, but the users still have to additionally parse the XML output from the service.

Given that the GeoNames gazetteer contains over eight million geographic place names (http://www.geonames.org/), this project only considers places that are close to coastlines and that are presumably influenced by coastal geomorphologic processes. The states considered are Massachusetts, Florida, and Alaska.

In addition because the data source for the GeoNames gazetteer is large (its data source includes the CIA World Factbook, Wikipedia, and USGS GNIS to name just a new, see http://www.geonames.org/data-sources.html for the complete list), a search entry to the service with place name, county name and state name could result in a number of records. To constrain the problem the research only retrieves the first record assuming that the first result usually is the closest to the search.

3.3.2.3 Matching Geographic Footprints with Polygons and Retrieving Topological Relationships

GeoTools (http://geotools.codehaus.org/) is chosen to process spatial data. GeoTools has a rich Java Class Library that allows applications written in Java to access and query spatial geometries stored in the Environmental Systems Research Institute (ESRI) Shapefile format. This factor is important in coordinating with the GeoNames Java Client that is also written in Java. To avoid issues resulting from using different coordinate systems; the geographic coordinates from the gazetteer are transformed to the coordinate system used in the spatial data by applying the Java Map Projection Library (http://www.jhlabs.com/java/maps/proj/).

Based on the spatial functions provided in GeoTools, the topological relationships between the candidate polygon and its surroundings are queried. The resulting topological relationships are then checked with the rules of conceptual neighborhoods and topological constraints developed in the previous steps.

3.3.2.4 Developing the Geographic Ontology in OWL

If the resulting topological relationships are the required neighboring relationships, the feature name is obtained and connected with the feature class of the place in the ontology, assuming that the concept does not yet exist in the ontology and that it is different from the feature class. Figure 3.10 shows the concept “lake” “evolves_into” “bay” in OWL.

```
<owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#lake">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:someValuesFrom>
        <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#bay"/>
      </owl:someValuesFrom>
      <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
    </owl:Restriction>
  </rdfs:subClassOf>
      .
      .
</owl:Class>
```

Figure 3.10: The ontology in OWL that describes “lake” evolves into “bay”.

3.3.2.5 Evaluating the Developed Ontology

To ensure this research achieves its objectives, the following two criteria should be examined in the evaluation:

1) The developed ontology should play a complementary role to the FTT. If the ontology is just another replication of the FTT, it will not support the main objective of this research to utilize multiple gazetteers, since the currently available FTT could be used instead.

2) The research claims that the ontology, which provides evolutionary information about geographic features, is needed in matching gazetteers. Using the ontology to match gazetteers should improve the matching results, i.e. the ontology should be able to match place names that otherwise cannot be matched because these place names are referred to different feature classes.

The following outlines an overview how the evaluation can be conducted to test the ontology. The detail of the evaluation will be explained further in Chapter 5.

As the purpose of the evaluation is to compare the FTT and the ontology, only places that have the same name but are referred to as different feature classes in the gazetteers are considered in the evaluation. To achieve this, one local source is evaluated at a time. Because this research only focuses on natural hydrographic features, the place names from this source that do not belong to this category are removed. The resulting place names are then checked for their appearance in a global gazetteer (i.e. the GeoNames gazetteer) and if they are referred to as different feature classes. The final list contains only place names that appear and are referred to different feature classes in both sources. Because the place names involved can contain thousands of records, the task is done computationally.

Table 3.1 demonstrates examples of place names that can be collected. The place names from the left and center columns are obtained from the GeoNames gazetteer and the Florida Geographic Data Library (FGDL) respectively. The right column lists the county and state of the place. The hydrographic features in the brackets are the feature classes of the respective places.

Table 3.1: Places that can be collected from the GeoNames gazetteer and the Florida Geographic Data Library.

	GeoNames	Florida Geographic Data Library	County, State
(1)	Juniper Creek (Swamp)	Juniper Creek (STREAM)	Walton, FL
(2)	Vanns Slough (Inlet)	Vans Slough (GUT)	Volusia, FL
(3)	Double Pond Branch (Lake)	Double Pond Branch (STREAM)	Washington, FL
(4)	Kingsley Creek (Stream)	Kingsley Creek (GUT)	Nassau, FL

To match the resulting place names using the FTT and the ontology, a matching system is developed. Using this system, if the different feature classes are defined as related in either the FTT or the ontology, the place names from the sources are matched. As an example, assume that the ontology describes "stream" to evolve into "swamp" and "lake", and in the FTT, "stream" is related to "gut". Using the system, the first and third records will be matched by the ontology, and the fourth record will be matched by the FTT. Table 3.2 describes the summary of the matching result based on this assumption. As shown in the table, the ontology outperforms the FTT by 25%, and 75% of the categories that otherwise cannot be matched can be matched with the combination of the FTT and the ontology.

Table 3.2: The summary of the matching result using the FTT and the ontology.

		Percentage
Total number of mismatched categories	4	
Number of matches by the ontology	2	50%
Number of matches by the FTT	1	25%
Number of mismatches	1	25%

CHAPTER 4 RESULTS

4.1 Introduction

This chapter presents the results from the implementation phase reviewed in the previous chapter. Section 4.2 demonstrates the spatial data selected for this research along with a discussion of coordinate transformation to facilitate data compatibility. Section 4.3 shows the ontology development system and compares the developed ontology with the FTT. Section 4.4 explains how the FTT and the ontology can be used to match place names with the matching system. This matching system will then be used for the evaluations, which will be discussed in Chapter 5.

4.2 Data Source

The hydrography spatial data layer owned by the Office of Geographic and Environmental Information of Massachusetts (MassGIS) (http://www.mass.gov/mgis/massgis.htm) is selected in this research as the spatial data for developing the ontology. Online resources such as the Geographic Information Systems Laboratory at MIT (http://libraries.mit.edu/gis/data/datalinks/ statedataweb.html, accessed on March 15, 2009), and GISuser.com (http://www.gisuser.com/content/view/16966/, accessed on March 15, 2009) were used to determine the spatial data that can be obtained from the state GIS clearinghouse. This dataset, at the scale of 1:25,000, contains polygons that describe more than 15 hydrographic features. The dataset was also checked for its positional errors, and the result showed that it is highly acceptable. The metadata describing this dataset is located in Appendix A and was originally obtained from the MassGIS official website (http://www.mass.gov/mgis/hd.htm).

As this research only focuses on coastal features, inland hydrographic features, waterfalls for example, are removed beforehand in order to improve the query performance. Figure 4.1(a) illustrates a modified version of the hydrographic layer that is used in this research. In this dataset each polygon is associated with a specific code, which in turn represents a specific hydrographic feature in the real world. Figure 4.1(b) is a summary of the codes with their corresponding hydrographic features. The complete code list with respective features can be found in the metadata.

Figure 4.1(a): A modified version of the MassGIS Hydrography layer.

Code	*Feature*
111	Marsh or Wetland
115	Flats
116	Bay or Estuary or Gulf
400	Rapids
402	Falls
412	Stream
414	Ditch or Canal
415	Aqueduct
416	Flume
421	Lake or Pond
.	.

Figure 4.1(b): A summary of the codes with corresponding hydrographic features.

4.2.1 Coordinate Transformation

The place coordinates from the GeoNames gazetteer are in the World Geodetic System (WGS) 84 geographic coordinate system. This is different from the coordinate system for the hydrography layer, which is in the Massachusetts State Plane Mainland FIPS 2001 Coordinate System. In order to ensure that the place coordinates from the gazetteer are matched correctly with the spatial data, a coordinate transformation is necessary. Table 4.1 lists the parameters and values for the Massachusetts State Plane Mainland FIPS 2001 System. Before matching place coordinates on the spatial data, the ontology development system applies these values to transform the place coordinates.

Table 4.1: The coordinate transformation parameters and values for Massachusetts State Plane Mainland FIPS 2001.

Projection:	Lambert Conformal Conic
Central Meridian:	71.5 (decimal degrees)
Latitude of Origin:	41.0 (decimal degrees)
Standard Parallel #1:	71.717 (decimal degrees)
Standard Parallel #2:	42.683 (decimal degrees)
False Easting:	200000 meters
False Northing:	750000 meters
Datum:	North American Datum (NAD) 1983
Ellipsoid:	Geodetic Reference System (GRS) 1980

4.3 Ontology Development

4.3.1 The Initial Ontology

Figure 4.2(a-c) illustrates the initial ontology using Protégé (http://protege.stanford.edu/), an ontology editor developed at Stanford University. In particular, Figure 4.2(a) demonstrates that *stream* has a topological constraint *overlaps* with hydrographic features, and this topological constraint can relate to one or more hydrographic features (indicated by the red circle). Figure 4.2(b) shows *bay* also has an overlapping topological constraint, but this topological constraint can relate to only one hydrographic feature. Lastly Figure 4.2(c) shows that *lake* has a topological constraint *inside* with land regions or physiographic features.

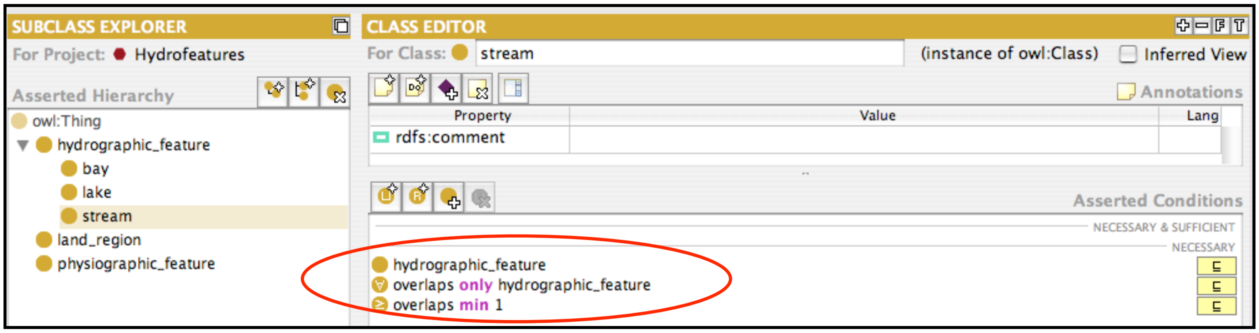


Figure 4.2(a): Topological constraints for *stream*.

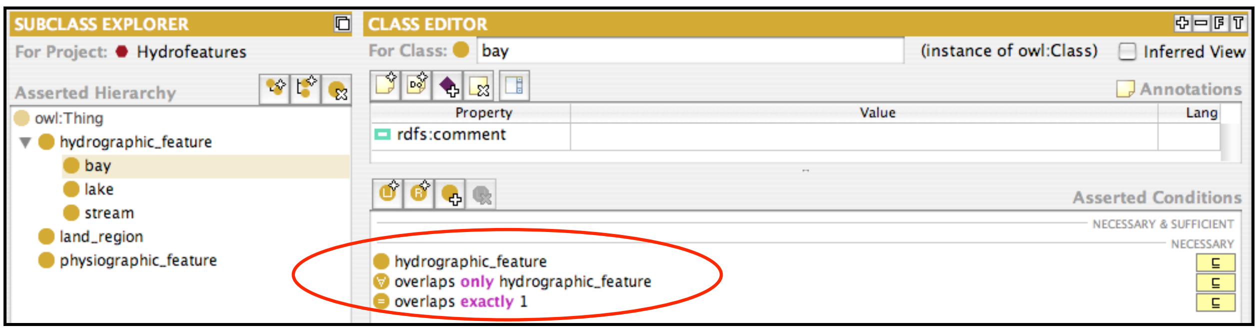


Figure 4.2(b): Topological constraints for *bay*.

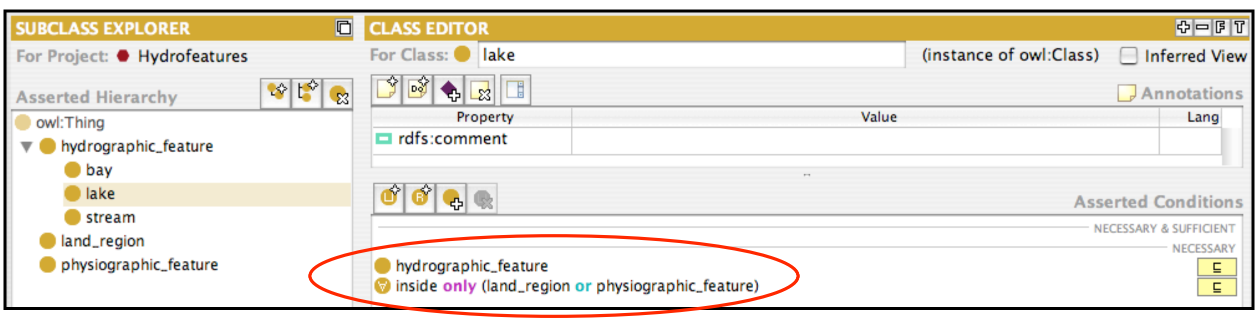


Figure 4.2(c): Topological constraints for *lake*.

Within each of the red circles, the symbol simply means the generalization relationship "is a" between concepts. For example in Figure 4.2(c), *lake* "is a" *hydrographic feature*. Below the generalization symbol is the universal operator . This operator signifies that *all* lakes are topologically located inside of a land region or a physiographic feature.

To provide a holistic view for the initial ontology, the ontology is represented in a graph structure depicted in Figure 4.3 using a Protégé plug-in called OWLViz (http://www.co-ode.org/downloads/owlviz/). As one can see from this graph structure, *stream*, *bay*, and *lake* have been related to *hydrographic feature* with the "is a" relation. This graph structure, however, also shows that, thus far, no evolutionary relationships have been defined between the concepts and no new concepts have been created from the spatial data.

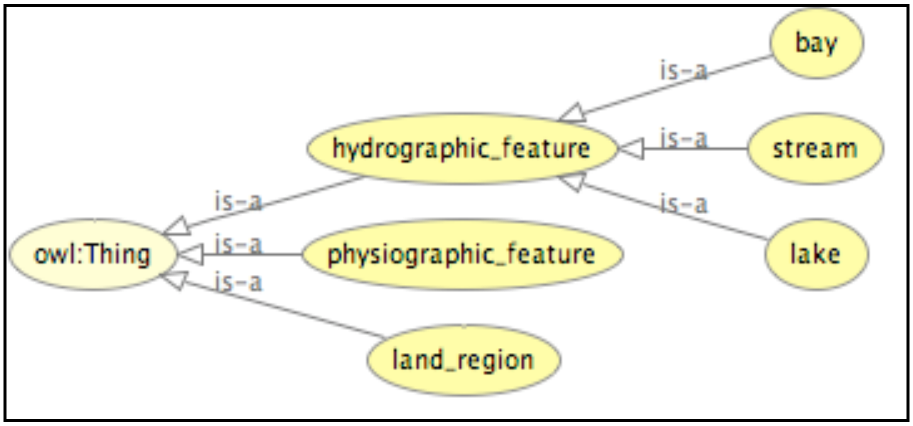


Figure 4.3: The initial ontology is represented in a graph structure using OWLViz.

4.3.2 The Ontology Development System

To define the evolutionary relationships between concepts and to create new concepts from the spatial data an ontology development system was created. Figure 4.4 illustrates the system interface. From this interface the user can load a text file containing place names by clicking on the "Browse" button at the top. This text file should be in the comma-delimited format. Each line in the file should contain place name, county name and state name with commas separating these names. Figure 4.5 shows an example of the format that the text file should follow. The place names are case-insensitive, meaning that "COW ISLAND POND", "Cow Island Pond", and "cow island pond" are read the same.

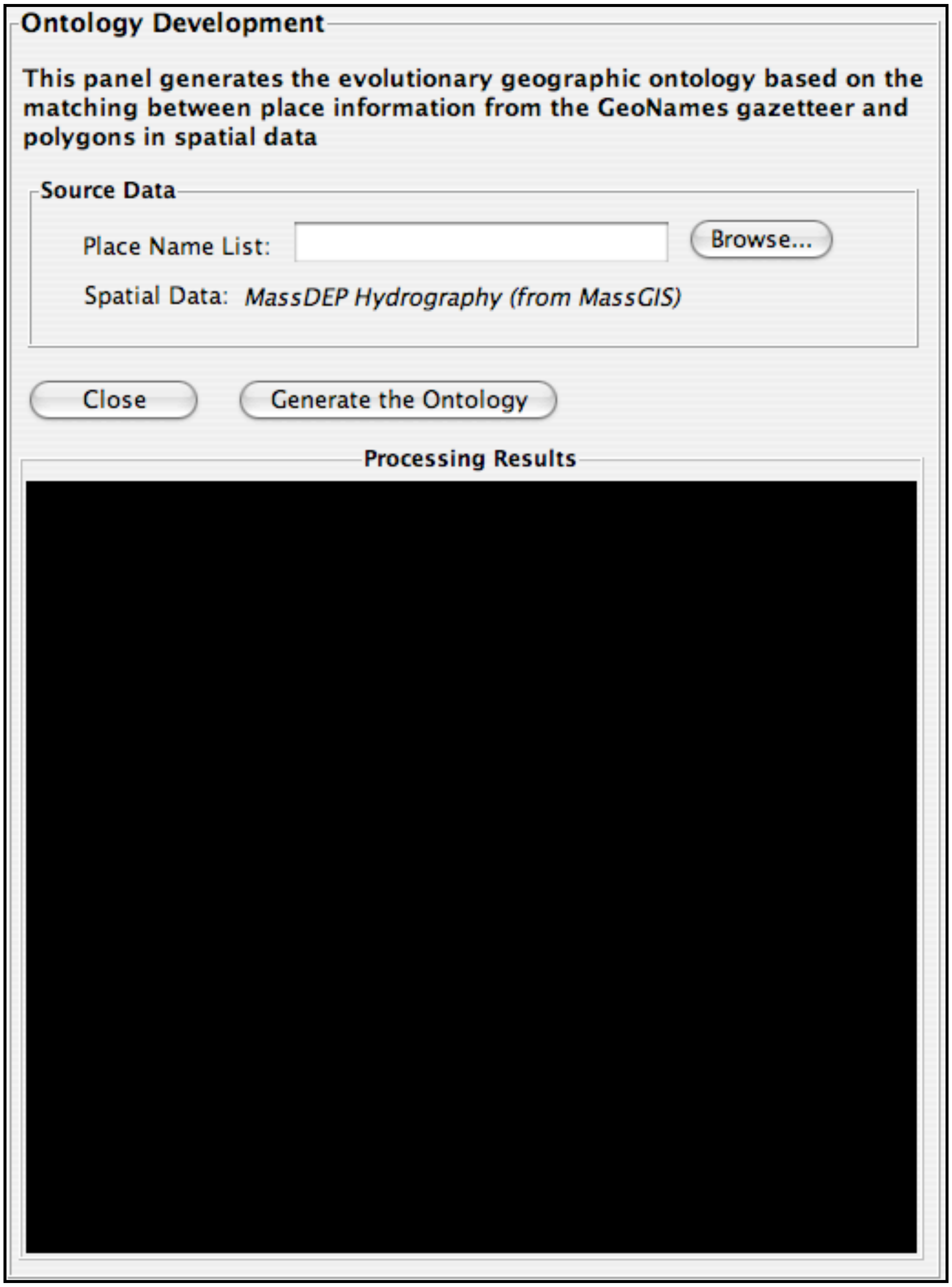


Figure 4.4: The interface of the ontology development system.

```
DOROTHY COVE, ESSEX,MA
LITTLE CEDAR SWAMP, BRISTOL,MA
COW ISLAND POND, NORFOLK,MA
NORTH RIVER, PLYMOUTH,MA
BROAD COVE, PLYMOUTH,MA
```

Figure 4.5: An example of the format that the input text file should follow to develop the ontology.

As an example, Figure 4.6 demonstrates a screenshot that shows processing descriptions after the user loaded the input text file and hit the command button “Generate the Ontology”. These processing descriptions inform the user on a step-by-step basis the processing steps taken by the system to create new concepts and evolutionary relationships in the ontology.

Specifically, when the user loads the list of place names, the system will first retrieve the place information containing the feature class, latitude and longitude for each of the place names from the gazetteer. As the gazetteer is a web service, this process is done in real time directly from the service.

After the place information is obtained, the system matches the footprint with the spatial data. If the footprint falls in one of the polygons, not only the feature name of the polygon is checked to ensure if it is different from the feature class, the topological relationships of the polygon should also meet the rules of low-resolution conceptual neighborhoods. As shown in Figure 4.6, the feature name, in this case “stream”, is different from the feature class (lake), and the topological constraint for the feature class is “inside”, the polygon should touch with other polygons. Given that the polygon has 4 touches with other polygons, it meets these requirements. The feature name is then checked with the ontology to see if it already exists. Since the feature name, stream, already exists in the ontology but not the evolutionary relationship between “lake” and “stream”, the relationship is created between the concepts.

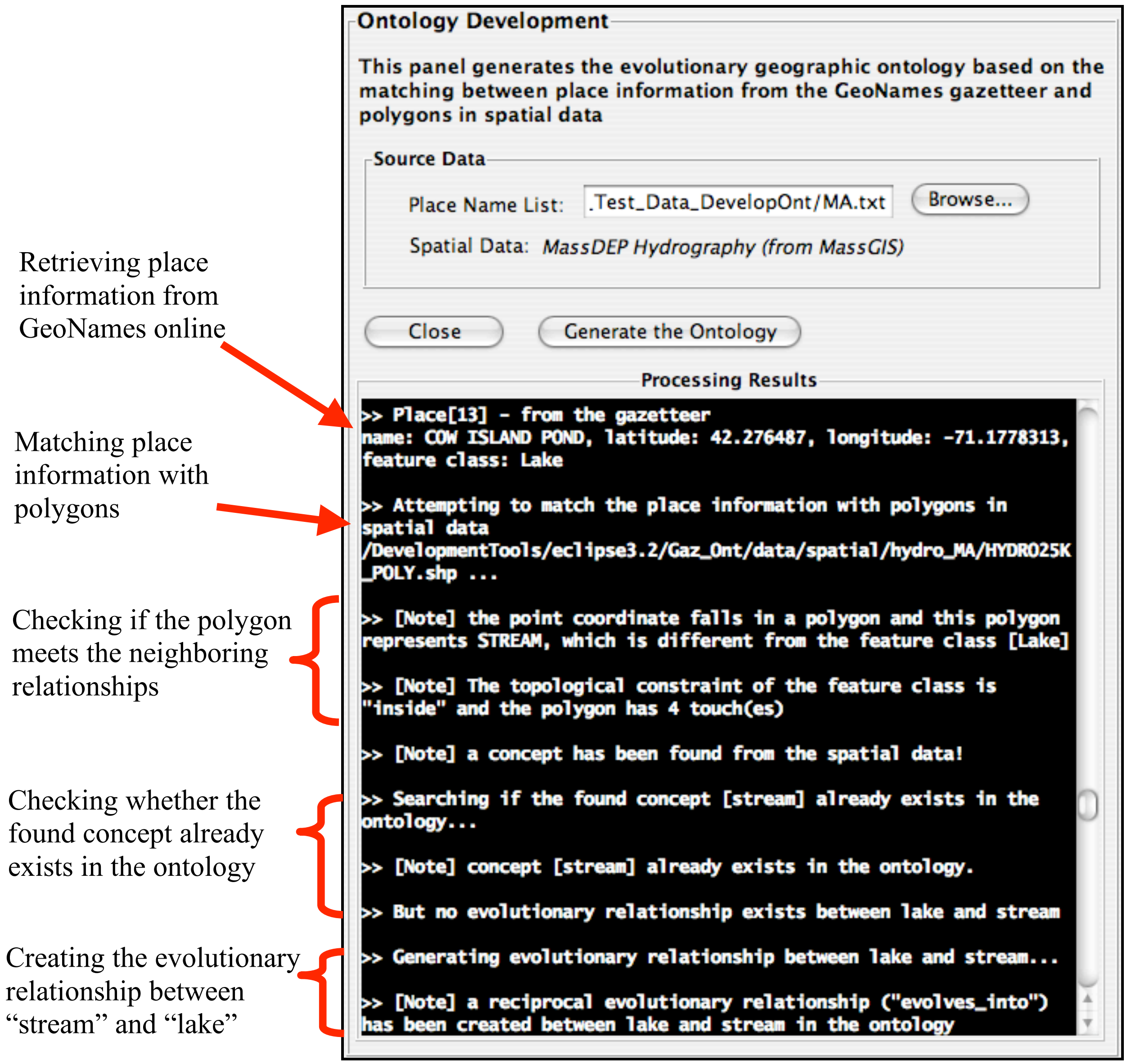


Figure 4.6: A screenshot shows processing descriptions being generated after the user loaded the text file containing place names and hit the "Generate the Ontology" command button.

4.3.3 The Evolutionary Geographic Ontology

Ultimately the ontology development system creates the evolutionary geographic ontology. The resulting ontology contains new concepts found from the spatial data and defines evolutionary relationships between concepts based on the matching and the rules of low-resolution conceptual neighborhoods. Figure 4.7 demonstrates the ontology developed from the system. Diverging from the initial ontology, this final ontology contains 9 new concepts, which are highlighted with red circles.

Besides adding new concepts, the system also creates evolutionary relationships between the concepts. Figure 4.8 illustrates an overview for the concepts that are related to the evolutionary relationships.

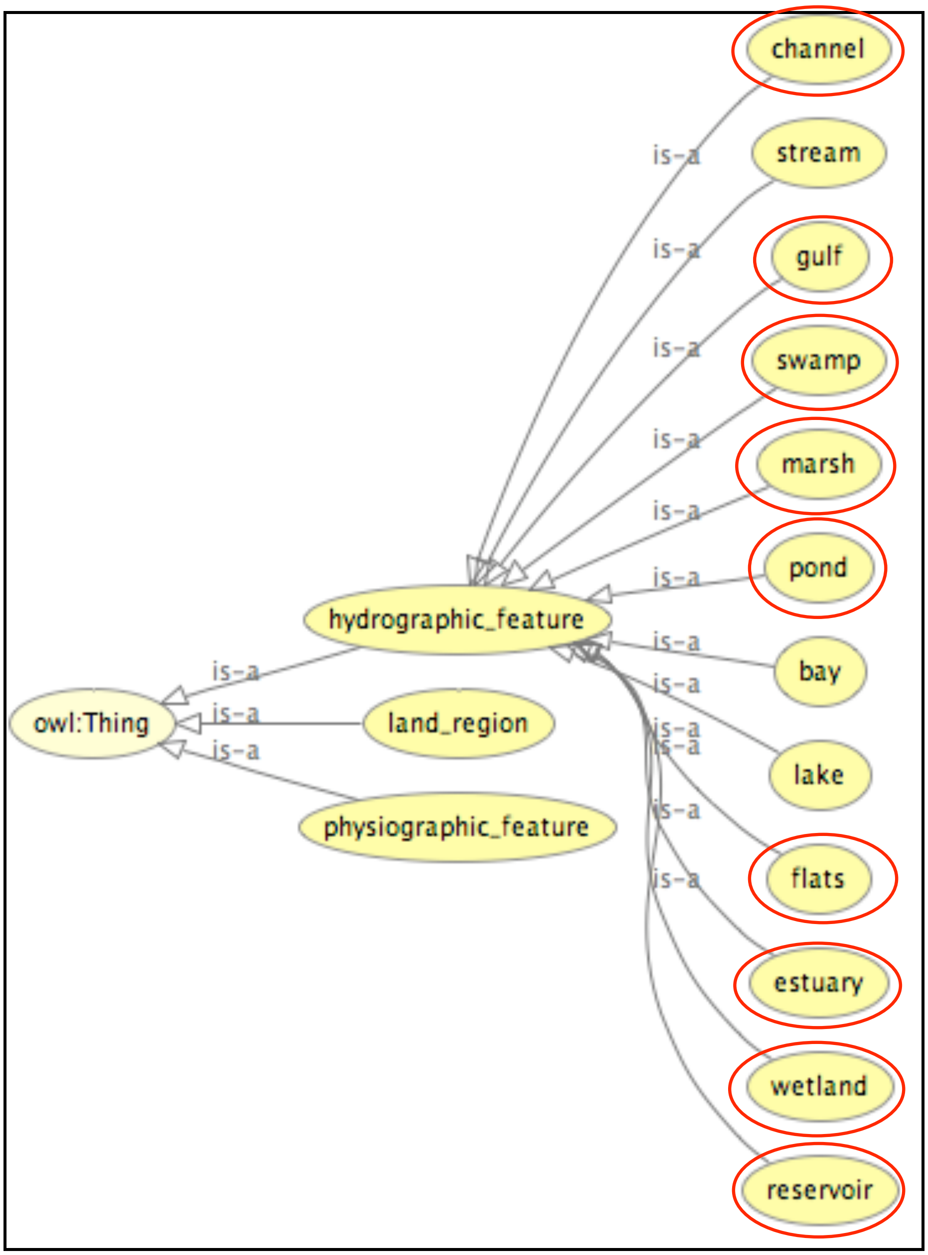


Figure 4.7: The final ontology that contains new concepts obtained from the spatial data.

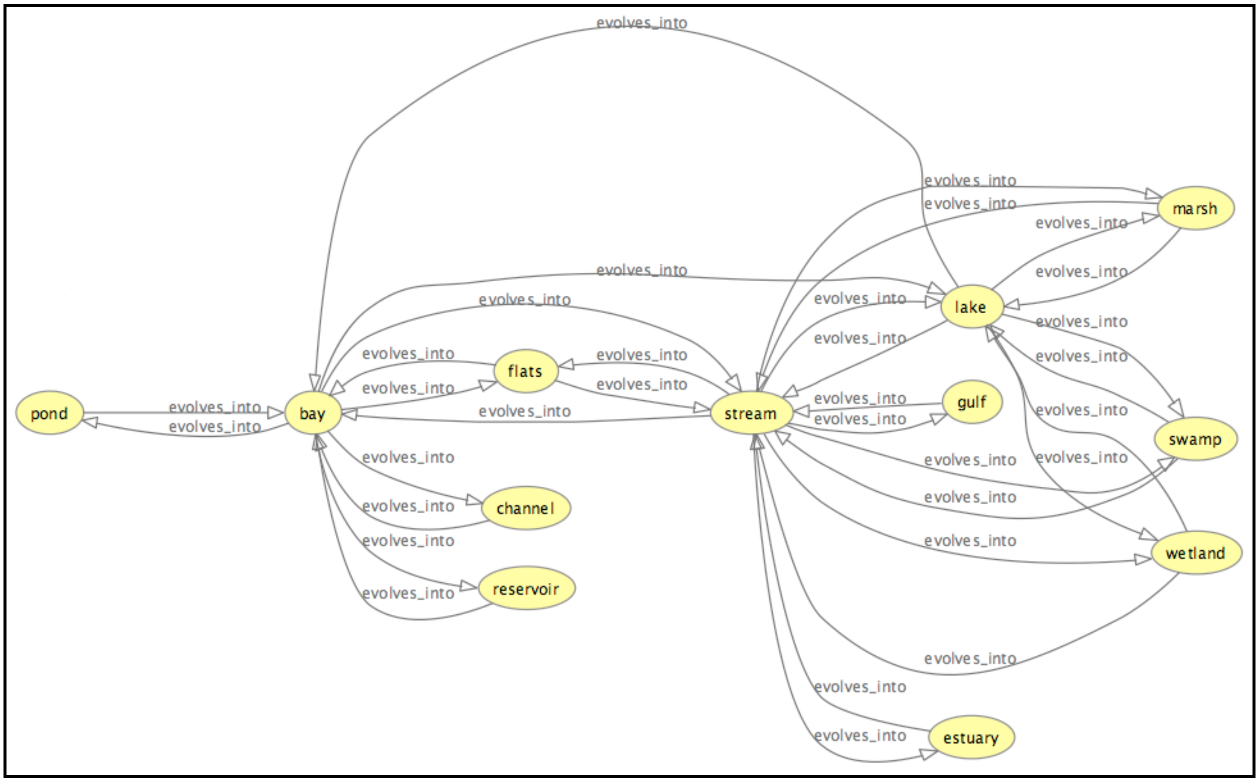


Figure 4.8: Concepts that are related to the evolutionary relationships in the final ontology.

Figure 4.9, Figure 4.10, and Figure 4.11 respectively show the updates to the three fundamental concepts - *stream*, *bay*, and *lake*. Compared to the initial ontology, these fundamental concepts have been defined with the evolutionary relationships in the final ontology. The symbol ∃ before the evolutionary relationships "evolve_into" indicates the existential operator. This existential operator indicates that *there exists at least one* stream that can evolve into an estuary (Figure 4.9).

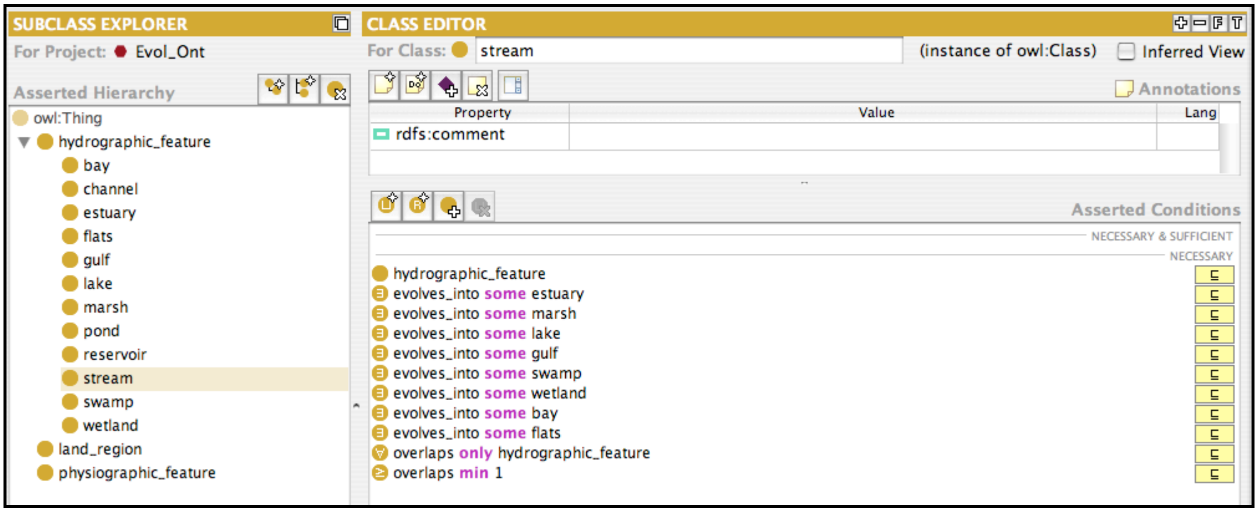


Figure 4.9: The concept *stream* has been updated with evolutionary relationships in relation to other hydrographic concepts.

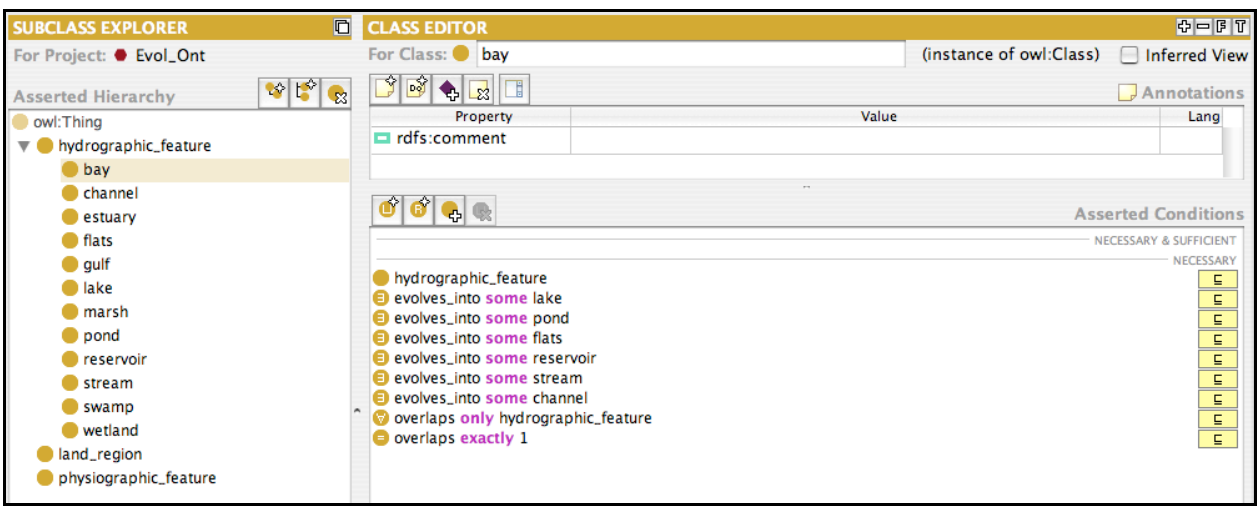


Figure 4.10: The concept *bay* has been updated with evolutionary relationships in relation to other hydrographic concepts.

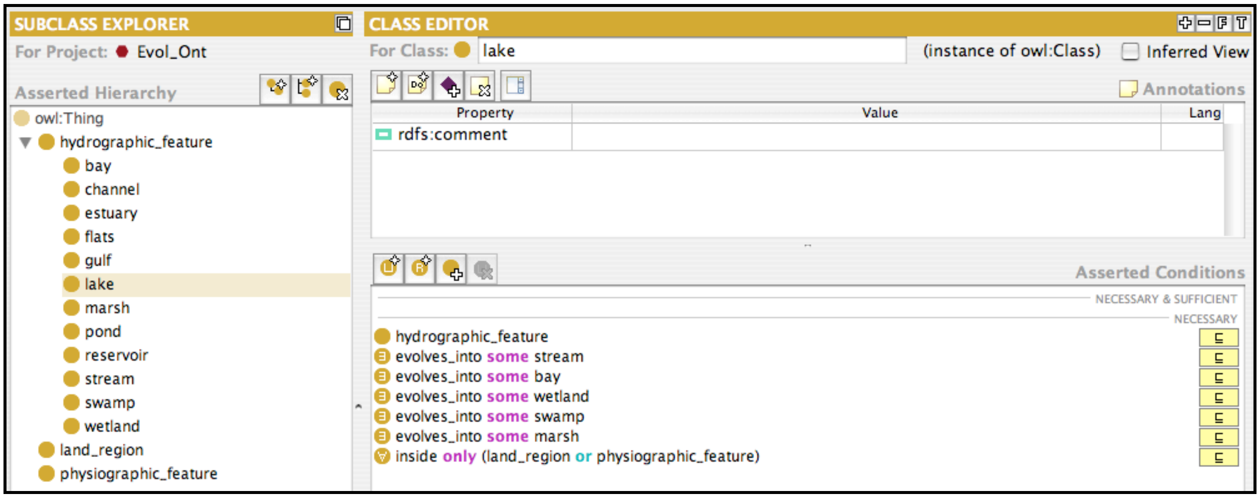


Figure 4.11: The concept *lake* has been updated with evolutionary relationships in relation to other hydrographic concepts.

4.3.3.1 Difference between the FTT and the Ontology

To demonstrate the difference between the FTT and the ontology, we limit our discussion to the three fundamental concepts. Figures 4.12 and 4.13 demonstrate the concepts according to the FTT and the ontology respectively. In Figure 4.12, the concepts that are located around each of the three fundamental concepts are related to the fundamental concept. For brevity, the relationships from these related concepts to the fundamental concept are omitted; only concepts that are shared by the three fundamental concepts are shown with relationships.

Comparing between the two figures, one can easily see that the FTT defines more concepts than the ontology. However a closer inspection of the concepts in Figure 4.12 reveals that these concepts are synonymous or specializations to the fundamental concept. For example with lake, pond and reservoir are synonymous, salt lake and crater lake are the specializations.

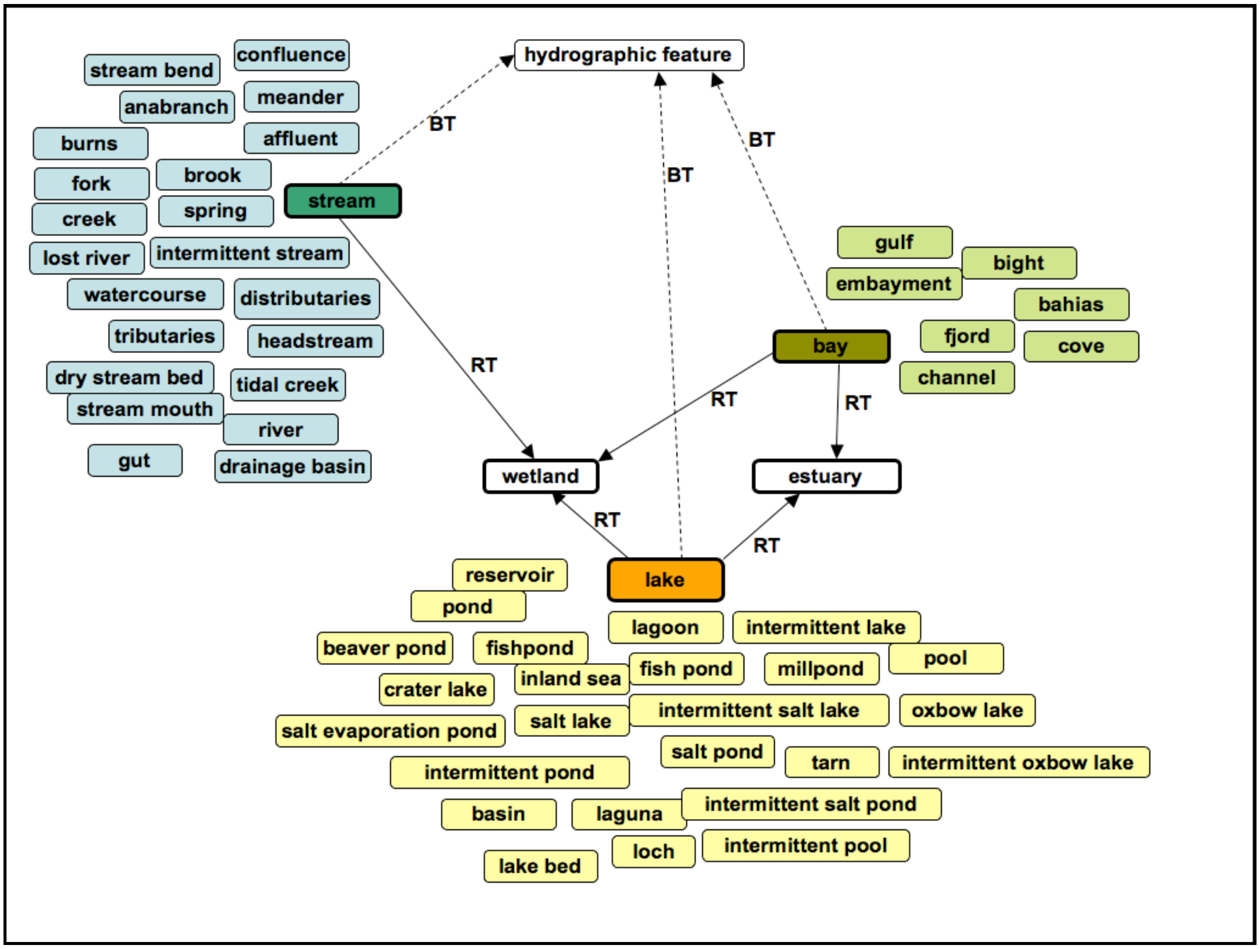


Figure 4.12: Concepts that are related with the three fundamental concepts in the FTT.

As opposed to the FTT, the concepts described in the ontology are related via topological changes between concepts. Examples are between stream and gulf, bay and channel, lake and bay, stream and lake. This research claims that such an ontology is crucial in matching gazetteers, given the fact that gazetteers document place names in different times. We will discuss the comparison between the FTT and the ontology in more detail in the next chapter when the matching results are produced based on the FTT and the ontology. In the following section, we will discuss how the FTT and the ontology will assist us in matching gazetteers using the matching system developed in this research.

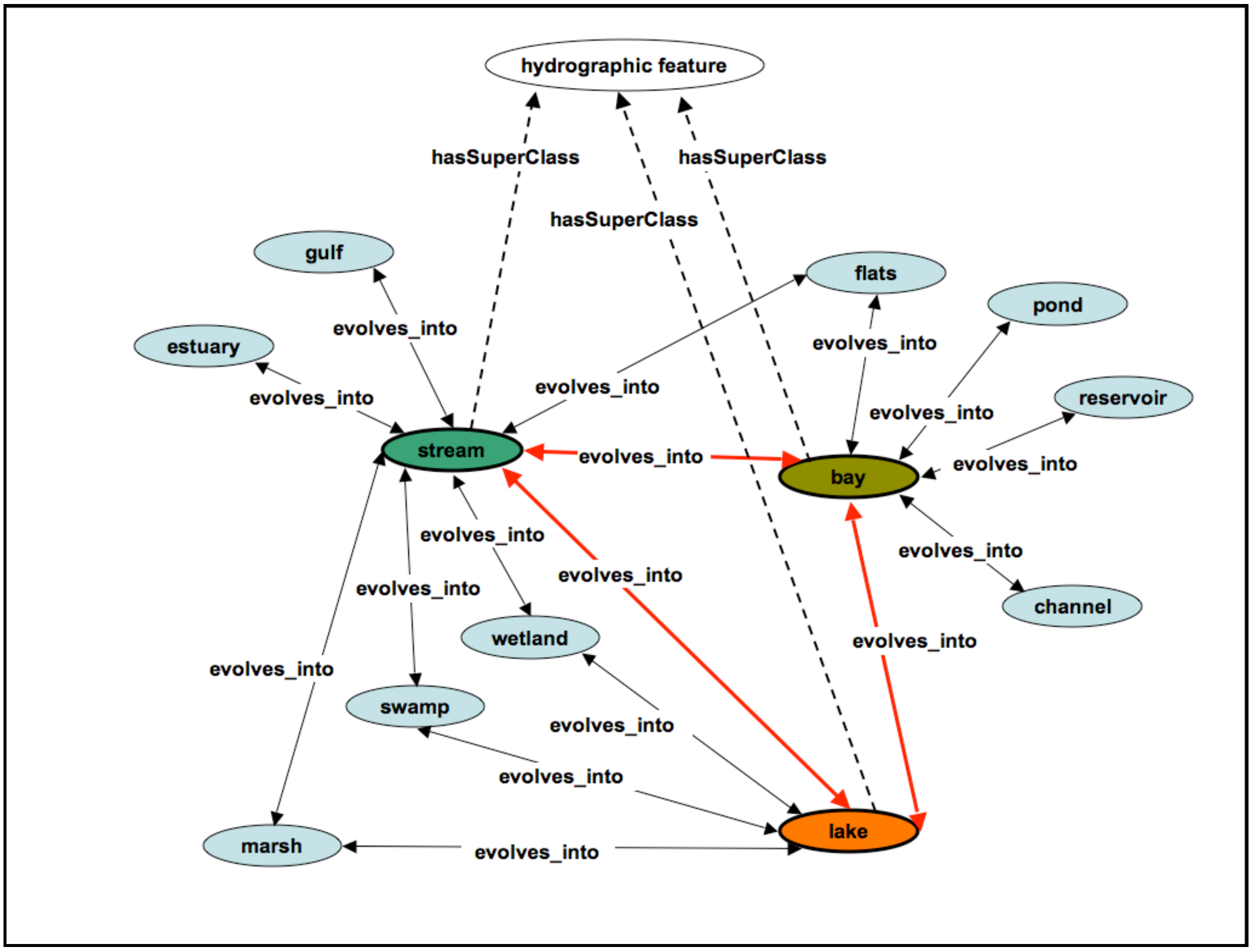


Figure 4.13: Concepts that are related with the fundamental concepts in the ontology.

4.4 Matching Gazetteers

The evolutionary geographic ontology can be represented as a graph structure as shown in the previous sections. The ontology may also be formatted in Web Ontology Language (OWL). In this research, the most significant advantage of representing the ontology in OWL is to facilitate matching between gazetteers. This section discusses how the FTT and the ontology are used for matching gazetteers. The matching system is then used to demonstrate matching gazetteers using the ontology in OWL and the original Feature Type Thesaurus (FTT), which is represented in eXtensible Markup Language (XML).

4.4.1 Matching Gazetteers using the Original ADL FTT and the Ontology

To illustrate how the FTT and the ontology are used for matching gazetteers, we focus on the example concept *stream* from the FTT and the ontology. Figure 4.14 describes the concept *stream* from the FTT in the XML format. In the FTT, dates the concept is inserted and updated are a part of the definition. Because this information is not relevant to our discussion here, it is omitted. Likewise, the definition of the concept *stream* from the ontology is also truncated. The topological constraints defined for the concept are not shown here for brevity. The full version of the ontology is attached in Appendix B for further reference.

To obtain related concepts from the FTT, the open source Java API for XML Processing (JAXP) supported by the Sun Microsystems (https://jaxp.dev.java.net/) was used to parse the FTT in XML. With additional coding plus the JAXP, the related concepts thus can readily be retrieved from the FTT. As an example, when searching for concepts that are related to *stream* on the FTT, concepts that are highlighted in yellow in Figure 4.14 will be returned.

```
<concept>
        <descriptor>streams</descriptor>
        <UF>affluents</UF>
        <UF>anabranches</UF>
        <UF>brooks</UF>
        <UF>burns (hydrographic)</UF>
        <UF>confluences</UF>
        <UF>creeks</UF>
        <UF>distributaries</UF>
        <UF>dry stream beds</UF>
        <UF>forks (physiographic features)</UF>
        <UF>headstreams</UF>
        <UF>intermittent streams</UF>
        <UF>lost rivers</UF>
        <UF>meanders</UF>
        <UF>stream bends</UF>
        <UF>stream mouths</UF>
        <UF>tidal creeks</UF>
        <UF>tributaries</UF>
        <UF>watercourses</UF>
        <BT>hydrographic features</BT>
        <NT>rivers</NT>
        <NT>springs (hydrographic)</NT>
        <RT>drainage basins</RT>
        <RT>guts</RT>
        <RT>wetlands</RT>
                .
                .
</concept>
```

Figure 4.14: The concept “stream” in the FTT.

For processing the ontology in OWL, as previously mentioned, the Jena API (http://jena.sourceforge.net/) is used. With some additional coding, the Jena API allows the system to retrieve related concepts from the ontology. For example in Figure 4.15, if the concept of interest is stream, the related concepts in blue can be retrieved from the ontology.

With the capability to retrieve related concepts from the FTT and the ontology, matching between gazetteers using the FTT and the ontology becomes possible. The following section demonstrates the matching system.

```
<owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#stream">
  <rdfs:subClassOf><owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf><owl:Restriction>
    <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
    <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#marsh"/></owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf><owl:Restriction>
    <owl:someValuesFrom><owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#wetland"/>
    </owl:someValuesFrom>
    <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/></owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf><owl:Restriction>
    <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
    <owl:someValuesFrom><owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#swamp"/>
    </owl:someValuesFrom></owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf><owl:Restriction>
    <owl:someValuesFrom><owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#lake"/>
    </owl:someValuesFrom>
    <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/></owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf><owl:Restriction>
    <owl:someValuesFrom><owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#estuary"/>
    </owl:someValuesFrom>
    <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/></owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf><owl:Restriction>
    <owl:someValuesFrom><owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#gulf"/>
    </owl:someValuesFrom>
    <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/></owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf><owl:Restriction>
    <owl:someValuesFrom><owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#flats"/>
    </owl:someValuesFrom>
    <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/></owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf><owl:Restriction>
    <owl:someValuesFrom><owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#bay"/>
    </owl:someValuesFrom>
    <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/></owl:Restriction>
  </rdfs:subClassOf>
      .
</owl:Class>
```

Figure 4.15: The concept “stream” in the ontology.

4.4.2 The Matching System

Figure 4.16 shows the interface of the matching system. To match place names from different gazetteers, the user first provides a list of place names where each place name indicates the names of the place and its county and state. The format of this list is the same as the format of the input text file for developing the ontology discussed in the previous section. To load the list, the user can click on the “Browse” button at the top of the interface.

Within the “Sources” box, the user is allowed to choose which two sources he is interested in using. As their place names cover different regions, the user can only choose between the Alaska State Geospatial-Data Clearinghouse and the Florida Geographic Data Library.

Before hitting the “Match” button, the user can choose between the FTT and the ontology or both to match the place names that have been loaded into the system. When the “Match” button is hit, the progress bar will be activated to show the progress of the matching. When the matching is completed, the results will be displayed as a table in the “Matching Result” panel. Alternatively, the user can also perform an exact string match between feature classes without selecting any available options. In such a case, neither the FTT nor the ontology will be considered. The system will only match place names that exactly have the same feature class in both sources. The research uses this option for the filtering process before the evaluation.

Generally when the “Match” button is hit, the matching system will look for every place name loaded by the user from the selected sources. If the place name is found in both sources, the feature classes that the place is referred to will be retrieved. If these feature classes are different, the system will retrieve related concepts from the FTT and/or the ontology. If the two feature classes are related according to the FTT and/or the ontology, the place names will be matched.

As an example, Figure 4.17 demonstrates the matching result after the user loaded the input file, made the selections, and hit the “Match” button. The matching results are displayed in four columns with the first two columns stating the sources of the place names, the third column specifying the county and state names of the place, and the last column informing whether the respective place names are matched based on the FTT and/or the ontology.

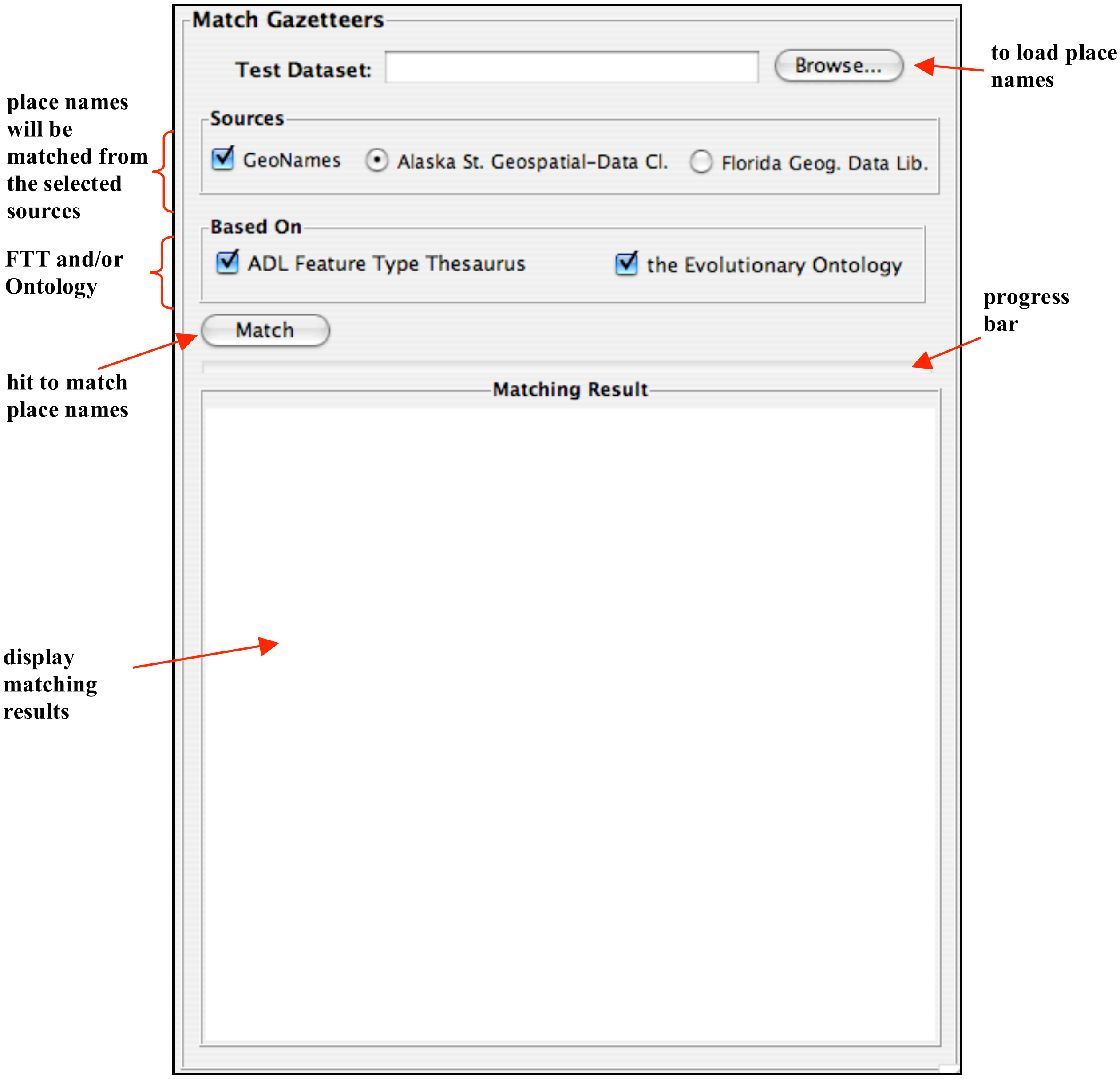


Figure 4.16: The interface of the matching system.

In the FTT, *stream* is related to *gut*, so the first record of the result is matched according to the FTT. From the ontology, as *stream* is related to *bay*, the second record is matched. Because no relation exists between *overfalls* and *falls*, in either the FTT or the ontology, the last record is not matched.

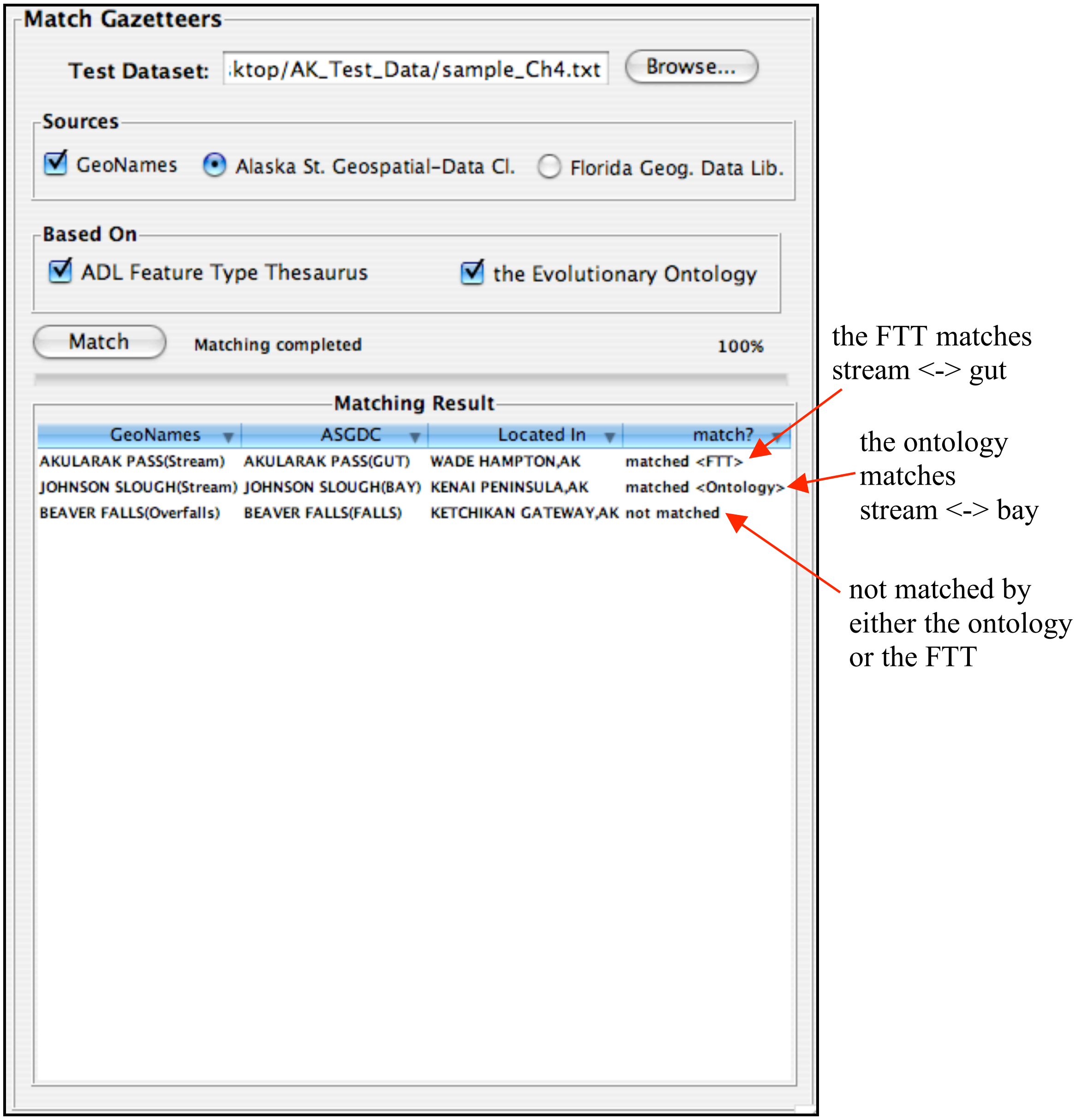


Figure 4.17: The result of matching based on the FTT and the ontology.

CHAPTER 5 EVALUATIONS AND DISCUSSIONS

5.1 Introduction

This chapter presents and discusses the matching results using the FTT and the ontology. Section 5.2 describes the test datasets and demonstrates the matching results. Section 5.3 compares and discusses these matching results in more detail.

5.2 Evaluations

This section is divided into two sub-sections, each of which corresponds to one evaluation. The first sub-section focuses on the evaluation between the Florida dataset and GeoNames, while the second focuses on the Alaska dataset and GeoNames. Although there are two evaluations, a few preliminary steps are necessary and common for both evaluations before the matching using the FTT and the ontology can be carried out. These steps are discussed here.

For each evaluation, all place names from a local dataset (place names from the Florida Geographic Data Library and Alaska Place Name Dictionary are considered as local datasets) are filtered beforehand, so that place names to be considered refer only to hydrographic features. The resulting place names are then checked with GeoNames to see if the names also appear and are referring to different feature classes. Generally this filtering process can result in the three groups of place names listed below:

1) Place names that appear in the local dataset do not appear in GeoNames. This is reasonable as the local dataset contains more specific place names, while the place names from GeoNames are more general given GeoNames focusing on the global context;

2) The same place names from the local dataset and GeoNames refer to *the same* feature class;

3) The same place names from the local dataset and GeoNames refer to *different* feature classes.

Of these three groups, we are only interested in the third group where the same place names from the local dataset and GeoNames refer to different feature classes. A difference between feature classes is then considered as one mismatched category provided one or more place names can belong to one of these categories. The performance evaluation of the FTT and the ontology is based on the production of the most mismatched categories. In the section below, the evaluations are discussed in more detail. The matching system discussed in the previous section is used to match the place names.

5.2.1 The Florida Dataset

The first test dataset is obtained from the Florida Geographic Data Library (FGDL). As stated in its official website (http://www.fgdl.org/), FGDL is "a mechanism for distributing spatial data (GIS) throughout the state of Florida" and it has participated in the Federal Geographic Data Committee (FGDC) Clearinghouse Registry.

The test dataset was downloaded from the FGDL FTP site (ftp://ftp1.fgdl.org/pub/state/). This site contains more than 350 datasets that cover different themes such as hydrography, soils, and transportation. All of them are compressed as zip files for ease of download. The dataset that is used for the evaluation is named "gnis_jul06.zip". It is a point-feature ESRI shapefile and supposedly contains all place names in Florida. According to the metadata from this dataset, these place names were originally obtained from USGS Geographic Names Information System (GNIS). The metadata of the dataset is attached in Appendix C.

The dataset initially contained 36,097 place names. With the filtering that removed place names that do not belong to hydrographic features, the remaining number of place names is 9,602. Figure 5.1 illustrates a screenshot of the filtered dataset in ESRI ArcExplorer. As one can see from the figure, each point feature includes name, county name, state name, and feature class. This information is required for matching place names with the GeoNames gazetteer.

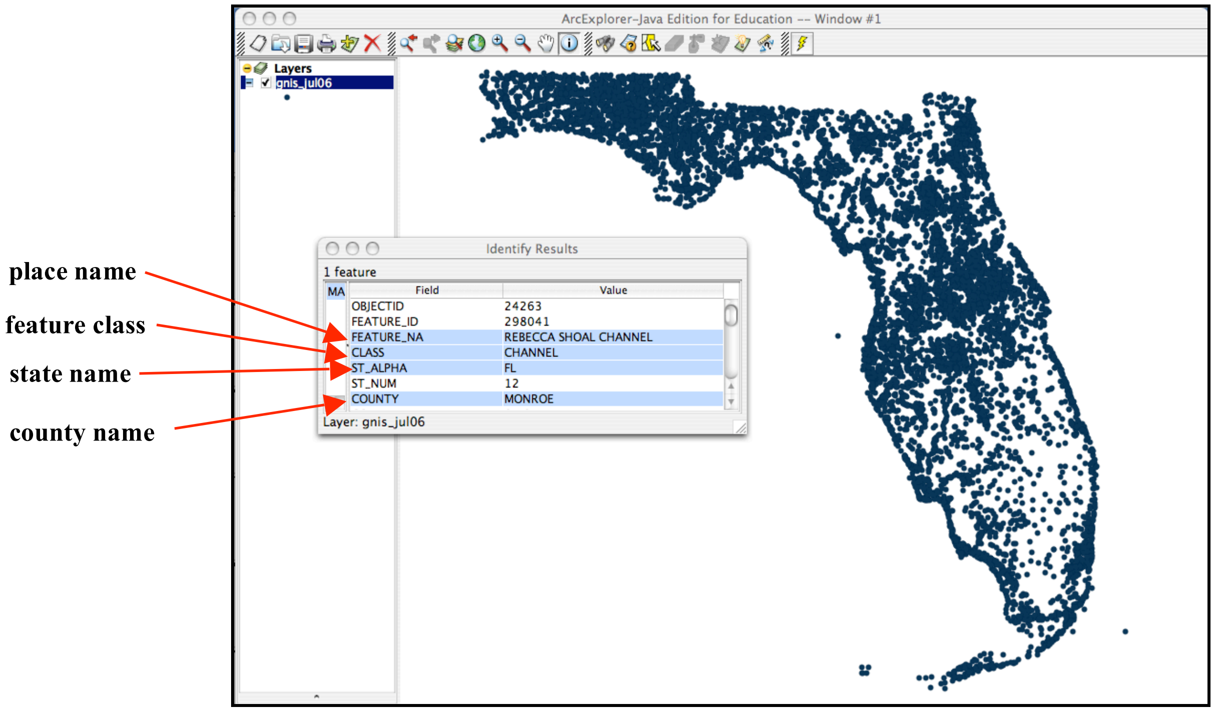


Figure 5.1: The Florida test dataset presented in ESRI ArcExplorer.

The remaining place names are then converted into the required format and are loaded into the matching system to check if 1) these place names from the Florida dataset also appear in GeoNames, and if they do, 2) the feature classes they belong are also the same or different. The result of this run is presented in Table 5.1.

Table 5.1: The information about the Florida dataset and the proportion of place names that have different feature classes between the dataset and GeoNames

		Percentage
Total number of hydrographic place names from the Florida dataset	9602	
Total number of hydrographic place names that also appear in GeoNames	9245	
Total number of place names that have different feature classes between the Florida dataset and GeoNames	457	457 / 9245 x 100 ≈ **5.0%**

The above result shows that using the same place name, which also includes the same county and state names, that about 5.0% of the total hydrographic place names that appear in both gazetteers will still refer to different feature classes.

457 place names that refer to different feature classes are then loaded into the matching system again, this time with the options "ADL Feature Type Thesaurus" and "the Evolutionary Ontology" being checked. The result of the matching by considering the FTT and the ontology is presented in Table 5.2 with the feature classes from the Florida dataset in uppercase.

As one can see from Table 5.2 all features in uppercase are hydrographic features. This is because all place names from the Florida dataset have been filtered out before the matching. In the case of GeoNames, the same place name can still refer to a land feature or an administrative unit, as no filtering has been made for GeoNames beforehand.

Because this evaluation considers only hydrographic features, mismatched categories with both feature classes that are not hydrographic features are removed. In addition, the feature classes appearing in the mismatched categories are not directional – i.e. feature class A mismatches with feature class B is the same as feature class B mismatches with feature class A, the mismatched categories therefore can be reduced to 16 categories. These final mismatched categories are presented in Table 5.3 as well as a bar chart with different color codes in Figure 5.2.

Table 5.2: The total mismatched categories between the Florida dataset and GeoNames.

Mismatched category	Number of place names belong the category	Matched?
Channel - BAY	2	matched <FTT>&<Ontology>
Lake - RESERVOIR	1	matched <FTT>
Reservoir - LAKE	2	matched <FTT>
Stream - GUT	9	matched <FTT>
Inlet - CHANNEL	2	matched <FTT>
Lake - STREAM	2	matched <Ontology>
Swamp - LAKE	3	matched <Ontology>
Bay - STREAM	5	matched <Ontology>
Stream - LAKE	2	matched <Ontology>
Swamp - STREAM	4	matched <Ontology>
Lake - BAY	1	matched <Ontology>
Lake - SWAMP	2	matched <Ontology>
Inlet - GUT	408	not matched

Valley - STREAM	2	not matched
Park - LAKE	1	not matched
Inlet - STREAM	2	not matched
Overfalls - FALLS	1	not matched
Stream - CHANNEL	1	not matched
Canal - STREAM	1	not matched
Cape - GUT	1	not matched
Island - SWAMP	2	not matched
Lake - GUT	1	not matched
Second-order administrative division - BAY	1	not matched
Swamp - GUT	1	not matched

Table 5.3: The mismatched categories that only contain hydrographic features.

Mismatched category	**Number of place names belong the category**	**Matched?**
Channel <-> Bay	2	matched <FTT>&<Ontology>
Lake <-> Reservoir	3	matched <FTT>
Stream <-> Gut	9	matched <FTT>
Inlet <-> Channel	2	matched <FTT>
Lake <-> Stream	4	matched <Ontology>
Swamp <-> Lake	5	matched <Ontology>
Bay <-> Stream	5	matched <Ontology>
Swamp <-> Stream	4	matched <Ontology>
Lake <-> Bay	1	matched <Ontology>
Inlet <-> Gut	408	not matched
Inlet <-> Stream	2	not matched
Overfalls <-> Falls	1	not matched
Stream <-> Channel	1	not matched
Canal <-> Stream	1	not matched
Lake <-> Gut	1	not matched
Swamp <-> Gut	1	not matched

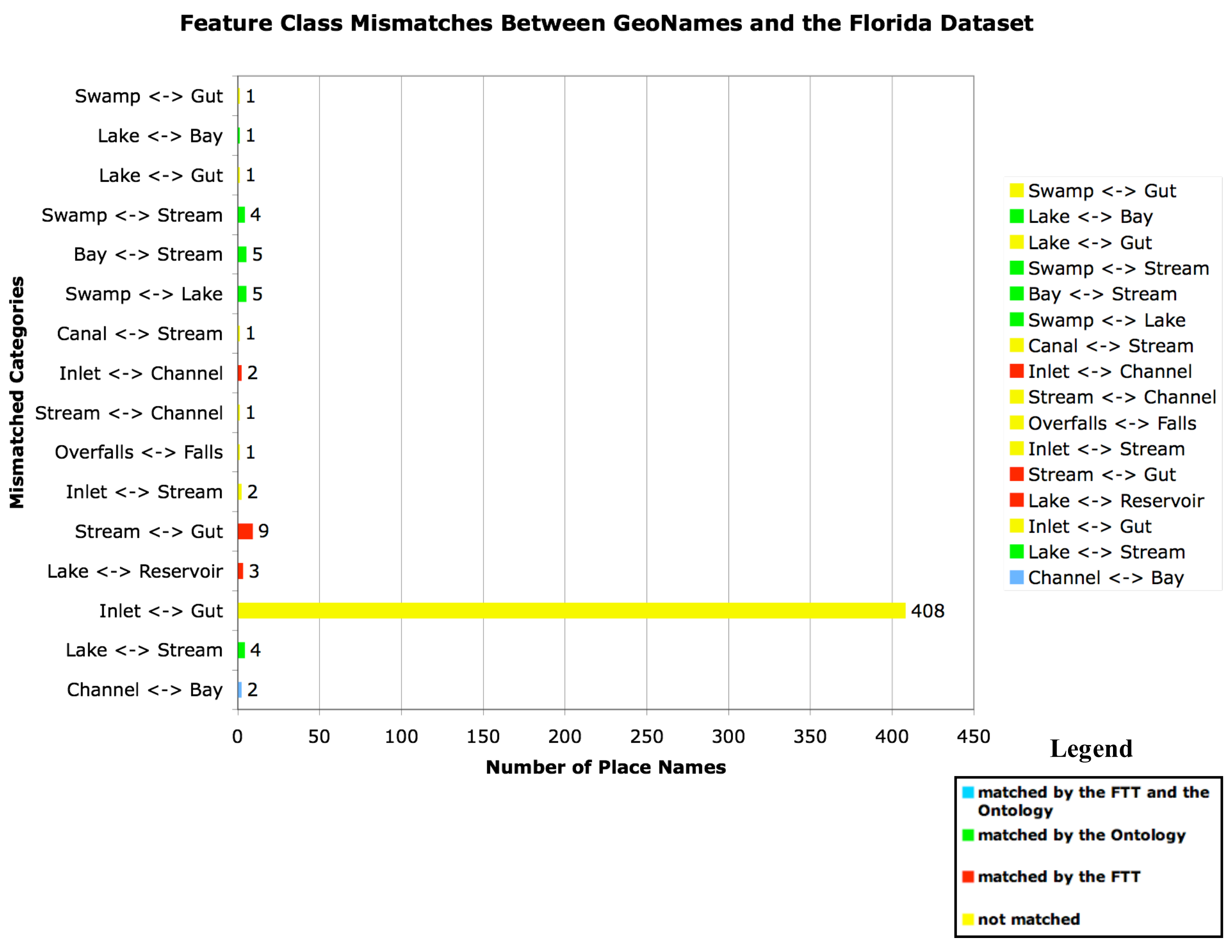


Figure 5.2: The bar chart for the mismatched hydrographic categories between the Florida dataset and GeoNames.

5.2.2 The Alaska Dataset

The second test dataset is obtained from the Alaska State Geospatial-Data Clearinghouse (ASGDC) (http://www.asgdc.state.ak.us/). Like FGDL, ASGDC provides free downloadable spatial data that cover different themes such as biology, culture, geoscience, and transportation.

The dataset used for testing was downloaded under the theme: "Cultural", and was named "Cultural – Alaska Place Names_POINT". Like the first dataset, this dataset is an ESRI shapefile with only point features. The place names in this dataset were originally obtained from the Alaska Place Name Dictionary (APND). The metadata describing this dataset is found in Appendix D.

The dataset initially had 25,876 place names, after the filtering, it was reduced to 13,088 place names, which are only hydrographic features. Figure 5.3 illustrates a screenshot of the Alaska dataset after the filtering. As one can see from the figure, each point feature is associated with name, feature class, and the county where the feature is located. As this is the dataset for place names in Alaska, we can also treat the state name for all the features in this dataset as Alaska.

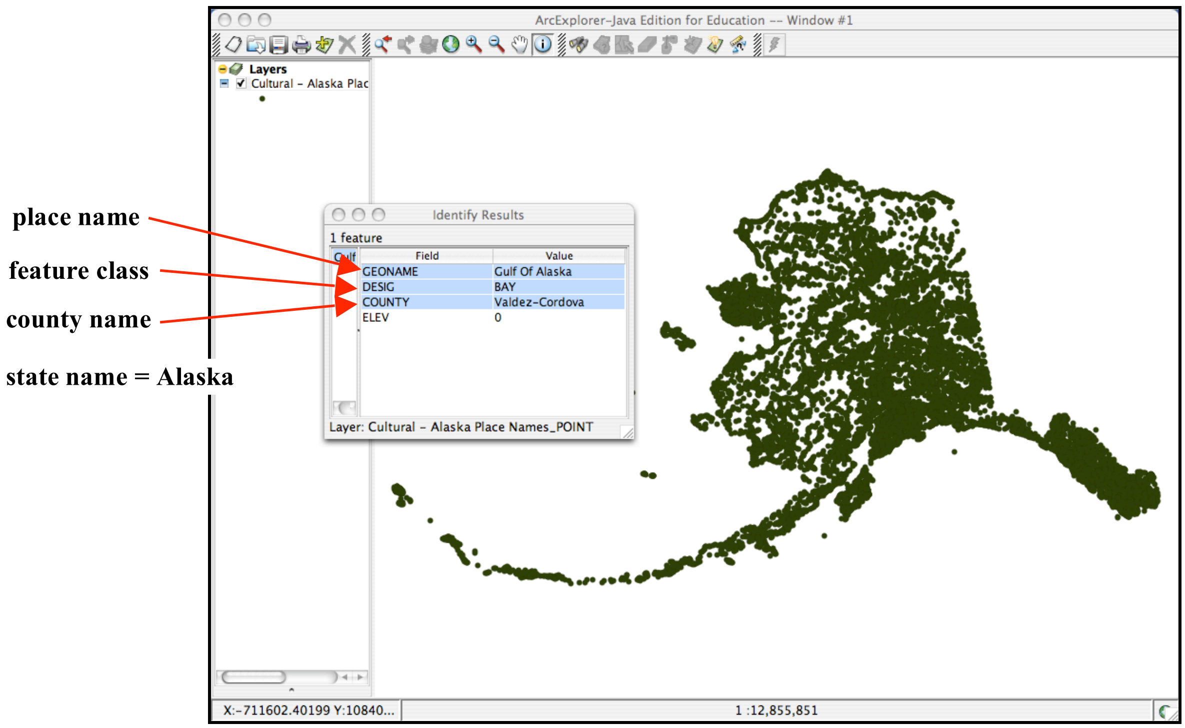


Figure 5.3: The Alaska dataset.

Table 5.4 shows the total number of place names that exist and refer to different feature classes between the Alaska dataset and GeoNames. Compared to the Florida dataset, this dataset has only about 2.1% of place names that have different feature classes.

Table 5.4: The information about the Alaska dataset and the proportion of place names that have different feature classes between the Alaska dataset and GeoNames.

		Percentage
Total number of hydrographic place names from the Alaska dataset	13088	
Total number of hydrographic place names that also appear in GeoNames	10823	
Total number of place names that have different feature classes between the Alaska dataset and GeoNames	223	223 / 10823 x 100 ≈ **2.1%**

Table 5.5 demonstrates the specific mismatched categories. Like the Florida dataset, these categories are composed of features that are not hydrographic in nature, for example airport, bar, and harbor. These mismatched categories are reduced and the mismatched categories that only contain hydrographic features are presented in Table 5.6 and as a bar chart in Figure 5.4.

Table 5.5: The total mismatched categories between the Alaska dataset and GeoNames.

Mismatched categories	Number of place names belong to the category	Matched?
Channel - BAY	1	matched <FTT>&<Ontology>
Lake - RESERVOIR	1	matched <FTT>
Stream - GUT	22	matched <FTT>
Channel - GUT	1	matched <FTT>
Reservoir - LAKE	6	matched <FTT>
Inlet - CHANNEL	9	matched <FTT>
Bay - LAKE	2	matched <Ontology>
Lake - STREAM	1	matched <Ontology>
Stream - BAY	1	matched <Ontology>
Bar - BAY	2	not matched
Airport - LAKE	3	not matched
Inlet - GUT	128	not matched
Airport - STREAM	1	not matched
Airport - BAY	1	not matched
Cape - BAY	1	not matched
Overfalls - FALLS	22	not matched
Harbor - BAY	13	not matched
Glacier - LAKE	1	not matched
Glacier - FALLS	4	not matched
Mountain - STREAM	1	not matched
Island - BAY	1	not matched
Glacier - STREAM	1	not matched

Table 5.6: The mismatched categories that are only limited to hydrographic features.

Mismatched categories	Number of place names belong to the category	Matched?
Channel <-> Bay	1	matched <FTT>&<Ontology>
Lake <-> Reservoir	7	matched <FTT>
Stream <-> Gut	22	matched <FTT>
Channel <-> Gut	1	matched <FTT>
Inlet <-> Channel	9	matched <FTT>
Bay <-> Lake	2	matched <Ontology>
Lake <-> Stream	1	matched <Ontology>
Stream <-> Bay	1	matched <Ontology>
Inlet <–> Gut	128	not matched
Overfalls <-> Falls	22	not matched
Glacier <-> Lake	1	not matched
Glacier <-> Falls	4	not matched
Glacier <-> Stream	1	not matched

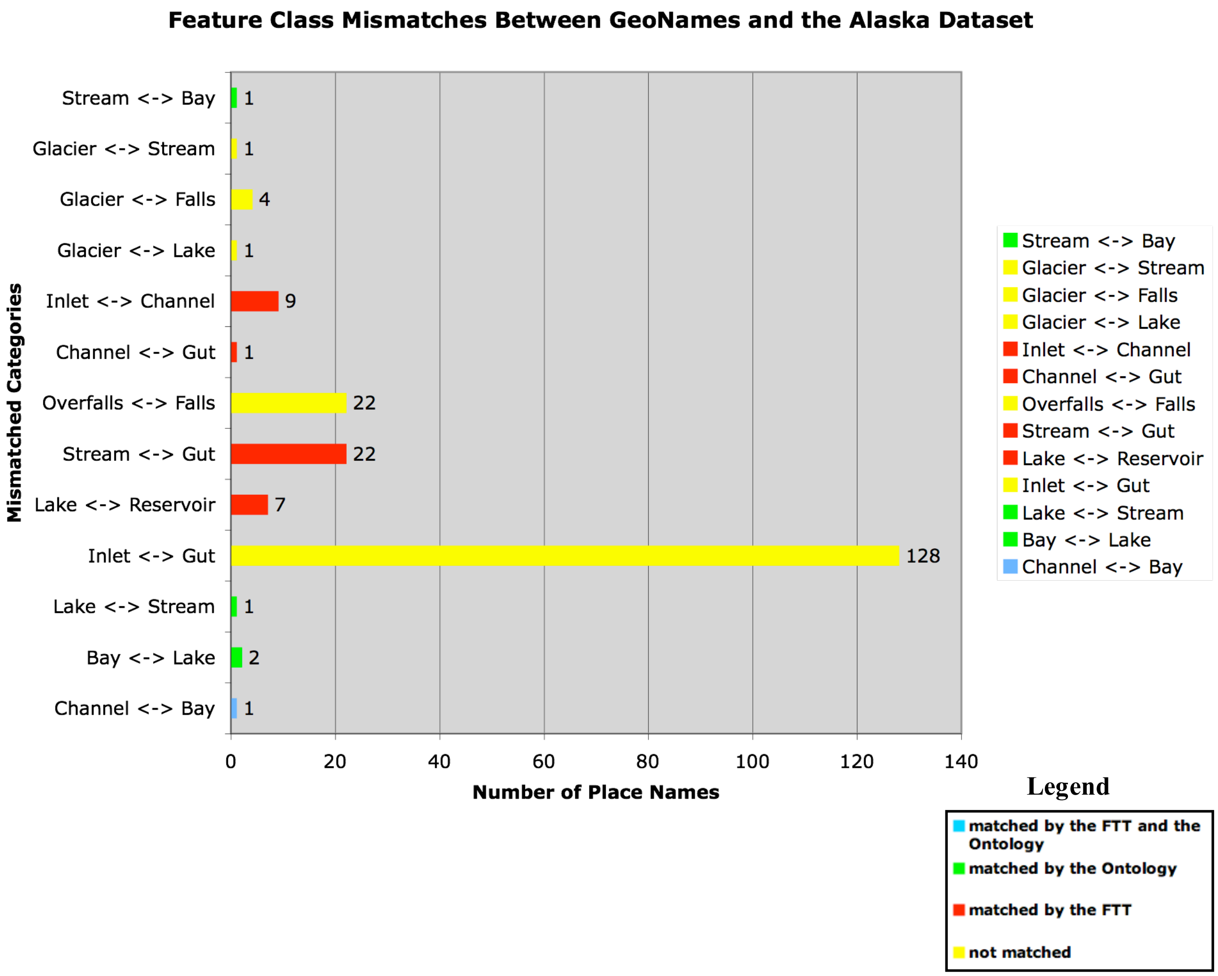


Figure 5.4: A bar chart presenting the mismatched hydrographic categories between the Alaska dataset and GeoNames.

5.3 Discussions

The summary of the matching result between GeoNames and the Florida dataset is described in Table 5.7. Table 5.7 shows that there are 16 mismatched categories in total. Of these 16 categories, the ontology matches 6 categories, which is 38%, compared to the FTT, which only matches 4 categories (25%). From this summary, we can say that the ontology outperforms the FTT in matching place names between GeoNames and the Florida dataset.

From this matching, we can also see that only one category can be matched by both the FTT and the ontology (i.e. "Channel <-> Bay"). This small overlap implies that the FTT and the ontology do not duplicate but complement each other. If the two are combined, the combination matches more than 50% of mismatched categories that otherwise cannot be matched.

Table 5.7: A summary of the matching result between GeoNames and the Florida dataset using the FTT and the ontology.

		Percentage
Total number of mismatched categories	16	
Number of matches by the ontology	6	38%
Number of matches by the FTT	4	25%
Number of mismatches	7	44%
Number of matches by the ontology AND the FTT	1	6%
Number of matches by the ontology AND/OR the FTT	9	56%

Table 5.8 presents the summary of the matching result between GeoNames and the Alaska dataset using the FTT and the ontology. As one can see from the table, of the 13 mismatched categories, the ontology can only match 4 or 31% and the FTT can match 5 or 38%. Although in this instance, the FTT slightly outperforms the ontology, if we combine this with the result from the Florida dataset the ontology generally slightly outperforms the FTT.

As in the Florida dataset, the combination of the FTT and the ontology can match more than 50% of mismatched categories between GeoNames and the Alaska dataset.

Table 5.8: A summary of the matching result for the Alaska dataset using the FTT and the ontology.

		Percentage
Total number of mismatched categories	13	
Number of matches by the ontology	4	31%
Number of matches by the FTT	5	38%
Number of mismatches	5	38%
Number of matches by the ontology AND the FTT	1	8%
Number of matches by the ontology AND/OR the FTT	8	62%

As a general summary, we can say that using the ontology alone to match place names for the datasets is slightly better than using the FTT alone. However if the FTT and the ontology are combined, they both can match more than 50% of the categories that otherwise cannot be matched.

In addition to comparing the FTT and the ontology, it is also important to see whether the two are complementary. Figure 5.5 and Figure 5.6 illustrate the matched categories by the FTT and the ontology respectively. From the list in Figure 5.5, except "Channel <-> Bay", all other respective feature classes can be considered the same topologically. For example channel and gut are elongated water bodies that *connect* with two other water bodies, and for lake and reservoir, they both are generally located *inside* of land feature.

In the FTT, "Channel" is related to "Bay" with "Related Term". As previously mentioned, if two terms are related according to one's view, one can relate the terms with "Related Term" in the FTT. "Related Term" is used to relate terms that cannot be related with either synonyms or a generalization relationship. If the developed ontology is to be integrated with the FTT for example, one can also consider "Related Term" to relate the concepts from the ontology to the FTT. This approach, however, will lose the specific relationship between the terms.

Instead of simply relating two terms/concepts with "Related Term", the developed ontology looks into topological change between terms. As shown in Figure 5.6, each matched category involves topological change between features (e.g. lake and stream), and this change is defined with the "evolve_into" relationship in the ontology. In other words, the ontology can play a complementary role with the FTT by enriching the relationship between concepts that are otherwise related with "Related Term".

From the evaluations in this chapter, the following three points can be made:

1. The FTT and the ontology individually represent different aspects of spatial knowledge. The FTT focuses on generalization knowledge and synonyms, while the ontology emphasizes evolutionary knowledge. They are complementary.

2. By looking into each of them alone in matching the datasets, the ontology slightly outperforms the FTT.

3. If the FTT and the ontology are combined, more than 50% of the mismatched categories can be matched.

Channel <-> Bay
Channel <-> Gut
Lake <-> Reservoir
Stream <-> Gut
Inlet <-> Channel

Figure 5.5: The categories that can be matched by the FTT.

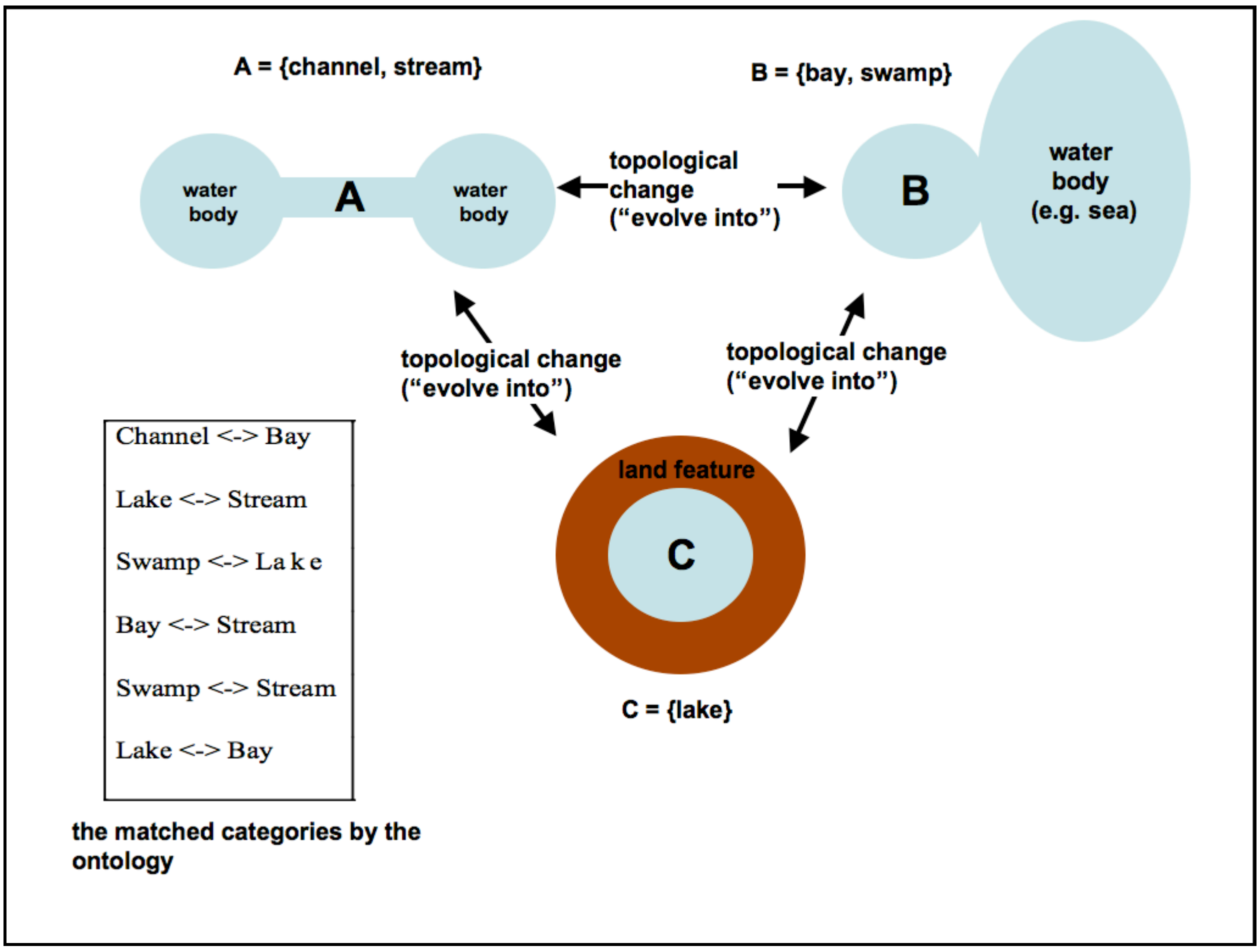


Figure 5.6: The categories that can be matched by the ontology. Each of the categories signifies topological change between features.

CHAPTER 6 CONCLUSIONS

6.1 General Conclusions

This thesis has focused on topological change between hydrographic concepts to develop a geographic ontology for matching gazetteers. As gazetteers recorded place names at different times, the same place may be referred to different geographic features due to changes in the environment. The goal of developing the ontology is to provide information about the possible changes between geographic features, so that when a feature evolves into a different feature, a computer system can use this ontology to match the gazetteers. Using the two local datasets from Florida and Alaska, the evaluations show that the developed ontology can possibly improve the matching between place names.

6.1.1 Objectives

Five objectives were addressed in this thesis. The first objective characterizes the topological constraints for hydrographic features. The second objective formalizes the evolution between hydrographic concepts using the notion of low-resolution conceptual neighborhoods. The third objective aims to match geographic footprints from the gazetteer with polygons in spatial data. The fourth objective is to develop the ontology. The fifth objective focuses on the evaluations between the FTT and the ontology.

The first and second objectives on characterizing the topological constraints and formalizing the evolutions between hydrographic concepts are met in Chapter 3 by presenting the conceptualizations of the fundamental concepts with respect to their topological nature. These conceptualizations allow for describing the concepts in Web Ontology Language (OWL) in Chapter 4, which in turn allows for the matching between geographic footprints and the spatial data. The first and second objectives should be integrated before the matching with the spatial data can be performed, and the topological constraints of a place should be known before the rules of conceptual neighborhoods are triggered.

The third objective, focused on the matching between footprints and spatial data, is described in Chapter 3 and implemented in Chapter 4. The initial testing by matching the selected footprints with spatial data manually in Chapter 3 prepares the computational implementation in Chapter 4.

Generally speaking, the first and second objectives look into the conceptual aspect of geographic features, while the third objective deals with the spatial aspect. When a feature meets these two aspects, it will be added into the ontology if not already present.

Developing the ontology is the fourth objective, and is addressed in Chapter 4. Most current applications dealing with ontologies tend to create ontologies manually. Automatically developing the ontology as shown in this thesis is a step forward in ontology development research.

The fifth and last objective, which is met in Chapter 5, sets out the evaluations between the FTT and the ontology. These evaluations show that not only does the ontology slightly outperform the FTT, but they are also complementary to each other in terms of represented knowledge. By combining the FTT and the ontology, more than 50% of mismatched feature classes can possibly be matched, as shown in the matching results.

6.2 Contributions

This section discusses the contributions of the thesis to GIScience, particularly in using multiple gazetteers and developing an intelligent GIS for sea level rise warning.

Figure 6.1 depicts a workflow for how the ontology (the “Geographic Ontology”) can be applied in using multiple gazetteers. As illustrated in Figure 6.1, if two corresponding places from the gazetteers refer to different feature classes, the ontology can be used to check if the involved feature classes have an evolutionary relationship. If a relationship exists, the related places can be matched. This contribution supports the nature of gazetteers that may document places in the past. The feature classes of past places may have evolved into different features at the current state, e.g. some places in New Orleans after the Hurricane Katrina, but the gazetteers may still refer to the places using previous feature classes. Using the ontology, old gazetteers can be used in conjunction with current gazetteers, even though the respective geographic features in the real world have changed. Being able to use multiple gazetteers is important for activities like urban planning where the information about places in the past is also required. Only by combining old and current place information, can one see the transformation of a region over time.

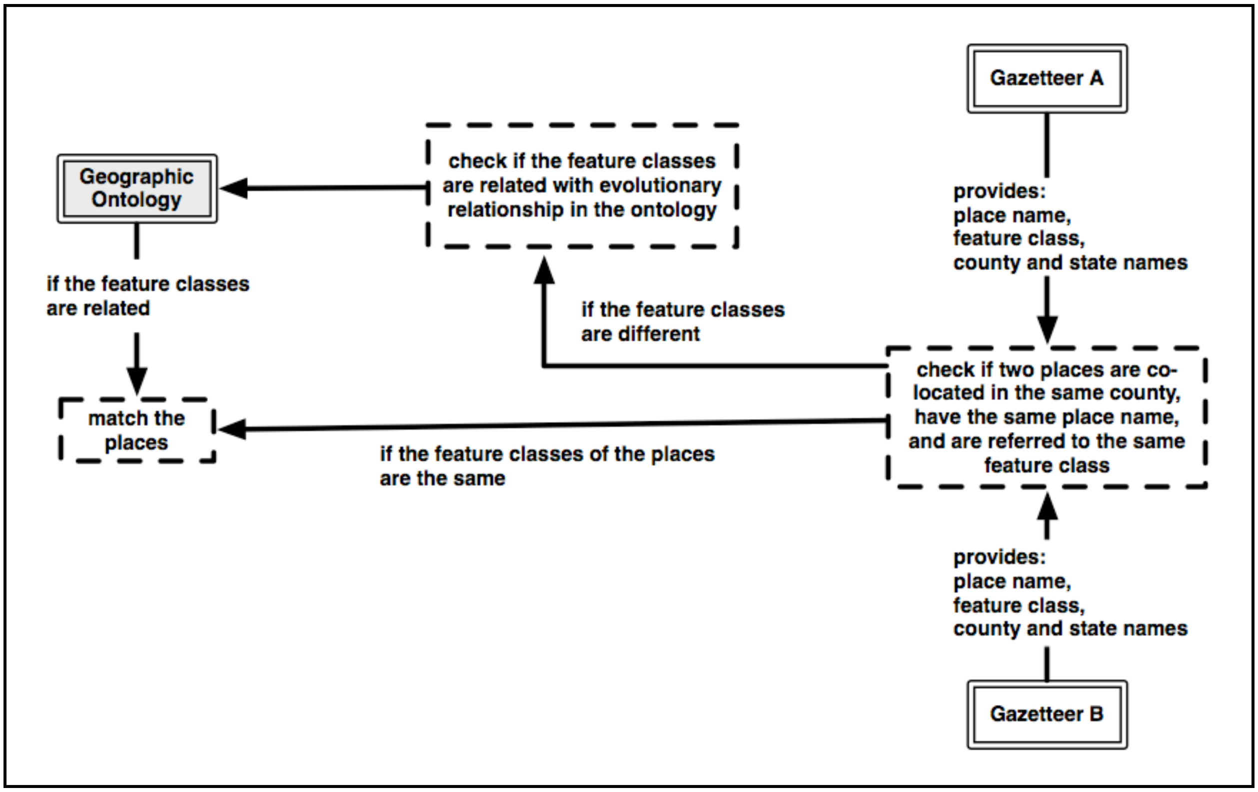


Figure 6.1: A workflow demonstrating that the developed ontology can be used to match multiple gazetteers.

Not only can the ontology facilitate utilization of multiple gazetteers, the ontology can also be used to study the phenomenological change of a region. As an example, the following describes a scenario where the ontology could be used to understand the issues of global warming in a GIS.

In the scenario, assume that the GIS is used as a tool to alert a community whenever the sea level rises. This GIS contains raster maps that identify geographic features in different colors, such as bays that are colored blue and lakes green. In order to be able to provide the most up-to-date information, the GIS updates these raster maps consistently every month from near-real-time remote sensing images. For each update, a classification method is performed to reclassify all features appearing on the maps. When the color of a particular region has changed after the re-classification, this signifies that the region has evolved into different features. At this point the GIS can only inform the community if the feature representing a particular area has evolved, the GIS cannot tell if the change resulted from environmental process such as sea level rise.

The ontology that describes evolutionary relationships between coastal features may support the GIS in determining the impact of environmental processes. In the ontology, the concepts are described with certain characteristics. These characteristics also include the topological nature of a feature, whether the feature is either completely or partially surrounded by another type of landform. This ontology is linked with the GIS after the classification method.

With the ontology, the GIS is then able to comprehend the characteristics of the features. This ability is important as it allows the GIS to interpret the change in more detail. For example, consider the GIS already knows that a feature is a non-self-connected water feature (e.g. a bay) from the ontology. When this feature has evolved into a self-connected water feature (e.g. a lake), the GIS can possibly interpret this change as a consequence of sea level decrease.

To interpret the change the meaning of the feature and the topological characteristics of that feature (i.e. self-connected or non-self-connected) from the ontology are crucial. The above case can be interpreted differently if the GIS detected a non-self-connected land feature to be evolved into a self-connected land feature (e.g. a cape to an island). In that case, the sea level may have risen instead. The meanings and the topological characteristics of coastal features are described in the ontology developed for this thesis.

6.3 Limitations and Recommendations

6.3.1 Hydrographic Features Are Not Topological Only

This project only considers topology to conceptualize hydrographic features. Topology however is not the only relationship that can be used to describe geographic features, other spatial relationships such as proximity (e.g. near) can also hold for geographic features. For example, a floodplain is always located near a river (Schwering 2004). In such a case we can use *near* as an inherent spatial relationship to represent the relationship between a floodplain and the river. This is just one example. More studies are needed to investigate the spatial relationships that can possibly exist in human conceptualization of geographic features. The results from such investigations can then be added to the ontology, which currently only considers topology, in order to cover more spatial knowledge for representing geographic features.

6.3.2 Human Subject Experiments

In this thesis, the ontology is developed mainly from a computational framework. A question arises if this ontology relates to human understanding. Although the ontology has proven useful in matching place names, a validation from the human perspective seems necessary.

To achieve this, a human subject experiment can be carried out. As an example, in the experiment all concepts without the relationships between the concepts, included in the ontology can be presented to the human participants. These participants are then asked to relate the concepts based on an evolutionary relationship. The end results from the participants are then compared with the developed ontology. For other detailed examples related to human subject experiments, the reader is referred to Smith and Mark (1999).

6.3.3 Comprehensiveness of Spatial Data to Include Distinct Areas

As the spatial data used in this research concentrates on the coastal region of Massachusetts, hydrographic features that are naturally located inland or far north are left untouched. The ontology seems futile if these features are referred as feature classes by certain places. This shortcoming is shown in the matching for the Florida and Alaska datasets discussed previously. In these datasets, hydrographic features such as overfall, fall, and glacier are referred by the place names. Because these features are not included in the ontology, the respective place names cannot be matched (Figure 6.2).

To improve the matching, the spatial data should cover extensive areas and these areas preferably should contain distinct geographic features.

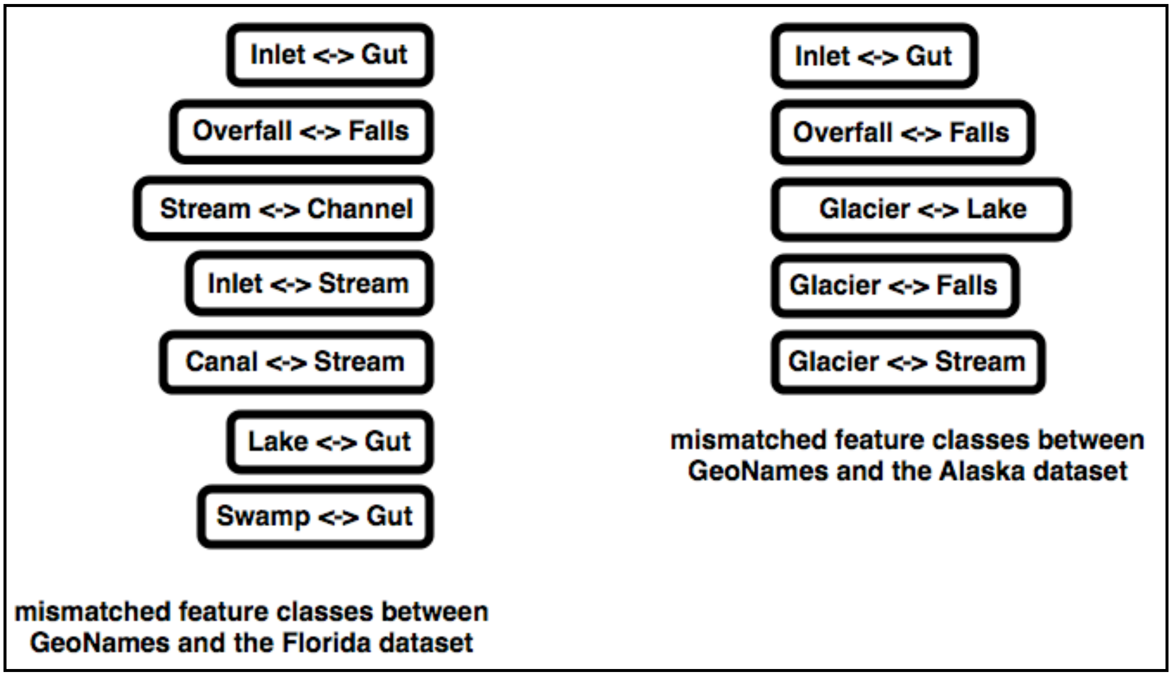


Figure 6.2: The mismatched feature classes in the Florida and Alaska datasets.

6.3.4 The Ontology Should Also Cover Land Features

Based on the framework, geographic features in the spatial data are required to be in polygon format. For features that are represented as linear features or point features, the framework cannot be implemented. This limitation eliminates land features such as mountains and capes, to be considered for developing the ontology.

Raster data representing land features may be used. Through a classification method, the "boundary" representing a given feature can be determined. In other words, in order to include land features in the ontology, raster data representing land features should first be vectorized to create "polygons", which allows for the system to check whether the geographic footprints fall in to one of these polygons.

REFERENCES

Bitters, B. 2005. A Geographical Ontology of Objects in the Visible Domain, College of Social Sciences, The Florida State University.

Burrough, P. A., and Frank, A. U. 1996. *Geographic Objects with Indeterminate Boundaries*: Taylor & Francis.

Carter, R. W. G., and Woodroffe, C. D. 1994. *Coastal Evolution: Late Quaternary Shoreline Morphodynamics*: Cambridge University Press.

Cohn, A.G., and Hazarika, S.M. 2001. Qualitative Spatial Representation and Reasoning: An Overview. *Fundamenta Informaticae* 43:2-32.

Cowell, P. J., and Thom, B. G. 1994. Morphodynamics of Coastal Evolution. *Coastal Evolution: Late Quaternary Shoreline Morphodynamics*:33–86.

Egenhofer, M. J., and Mark, D.M. 1995. Naive Geography. In *Spatial Information Theory: A Theoretical Basis for GIS*, 1-15. Semmering, Austria: Berlin: Springer-Verlag.

Egenhofer, M. J., and Al-Taha, K. K. 1992. Reasoning about Gradual Changes of Topological Relationships. *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*:196-219.

Fonseca, F., Camara, G., and Monteiro, A.M. 2006. A Framework for Measuring the Interoperability of Geo-Ontologies. *Spatial Cognition and Computation* 6 (4):309-331.

Goodchild, M. F., and Hill, L. L. 2008. Introduction to Digital Gazetteer Research. *International Journal of Geographical Information Science* 22 (10):1039 - 1044.

Hart, G., Mizen, H., and Temple, S. 2004. Tales of the River Bank: An overview of the First Stages in the Development of a Topographic Ontology. Paper read at 12th Annual Conference on GIS Research UK (GISRUK).

Hastings, J.T. 2008. Automated Conflation of Digital Gazetteer Data. *International Journal of Geographical Information Science* 22 (10):1109-1127.

Hill, L. 2006. *Georeferencing: The Geographic Associations of Information*: The MIT Press.

Hornsby, K., and Egenhofer, M.J. 2000. Identity-based Change: A Foundation for Spatio-Temporal Knowledge Representation. *International Journal of Geographical Information Science* 14 (3):207-224.

Janee, G. 2006. Rethinking Gazetteers and Interoperability. In *NCGIA Digital Gazetteer Research and Practice*. Santa Barbara, U.S.A: University of California at Santa Barbara.

Janowicz, K., and Keβler, C. 2008. The Role of Ontology in Improving Gazetteer Interaction. *International Journal of Geographical Information Science* 22 (10):1129 - 1157.

Lakhan, V. C. 2004. Perspectives on Special Issue Papers on Coastal Morphodynamic Modeling. *Coastal Engineering* 51 (8-9):657-660.

Leont'yev, I. O. 2003. Modeling the Morphological Response in A Coastal Zone for Different Temporal Scales. *Advances in Coastal Modeling*: 299-336.

Mark, D. M., and Turk, A. G. 2003. Landscape Categories in Yindjibarndi: Ontology, Environment, and Language. *Spatial Information Theory: Foundations of Geographic Information Science*:28–45.

Mizen, H., Dolbear, C., and Hart, G. 2005. Ontology Ontogeny: Understanding How an Ontology Is Created and Developed. *Lecture Notes in Computer Science* 3799:15.

Molenaar, M., and Cheng, T. 2000. Fuzzy Spatial Objects and Their Dynamics. *ISPRS Journal of Photogrammetry and Remote Sensing* 55 (3):164-175.

Purves, R.S., Clough, P., Jones, C.B., Arampatzis, A., Bucher, B., Finch, D., Fu, G., Joho, H., Syed, A.K., Vaid, S., and Yang, B. 2007. The Design and Implementation of SPIRIT: A Spatially Aware Search Engine for Information Retrieval on the Internet. *International Journal of Geographical Information Science* 21 (7):717-745.

Raper, J., and Livingstone, D. 1995. Development of A Geomorphological Spatial Model Using Object-Oriented Design. *International Journal of Geographical Information Science* 9 (4):359-383.

Smith, B., and Mark, D.M. 1999. Ontology with Human Subjects Testing: An Empirical Investigation of Geographic Categories. *American Journal of Economics and Sociology* 58:245-272.

Smith, B., and Mark, D. M. 1998. Ontology and Geographic Kinds. Paper read at International Symposium on Spatial Data Handling, at Vancouver, Canada.

Smith, D.A., and Crane, G. 2001. Disambiguating Geographic Names in a Historical Digital Library. Paper read at ECDI.

Schwering, A. 2004. Semantic Neighbourhood for Spatial Relations. 3rd International Conference on Geographic Information Science (GIScience 04), Adelphi, Maryland, October 20 - 23, 2004.

Third, A., Bennett, B., and Mallenby, D. 2007. Architecture for a Grounded Ontology of Geographic Information. *Lecture Notes In Computer Science* 4853:36.

J. Paul Getty Trust. 2008. *Getty Vocabulary Program Editorial Guidelines* 2008 [cited October, 20 2008]. Available from http://www.getty.edu/research/conducting_research/vocabularies/editorial_guidelines.html#tgn.

Volz, R. , Kleb, J., and Mueller, W. 2007. Towards Ontology-based Disambiguation of Geographical Identifiers. Paper read at WWW2007, at Banff, Canada.

Appendixes

Appendix A: Metadata for the MassGIS Hydrography Layer

The following data description was originally obtained from the official website of the Office of Geographic and Environmental Information (MassGIS) (http://www.mass.gov/mgis/hd.htm).

MassDEP Hydrography (1:25,000) - January 2009

OVERVIEW

Click to view larger size

The MassDEP Hydrography layer is an enhanced version of the older U.S. Geological survey 1:25,000 Hydrography datalayer. It represents hydrographic (water-related) features, including surface water (lakes, ponds, and reservoirs), wetlands, bogs, flats, rivers, streams, and others (see below).

The layer is a hybrid of data based on USGS Digital Line Graphs (DLGs), scanned mylar separates obtained from the USGS, digitized hydrographic features from paper USGS 1:25,000 Topographic Quadrangle maps and data extracted from the MassDEP Wetlands datalayer. Areas within many surface water supply watersheds have been enhanced by using higher resolution streams and lakes from the MassDEP Wetlands datalayer, many areas have also been field verified. This layer is intended as an interim product that will be incorporated into the USGS's National Hydrography Dataset (NHD) when funding and resources become available. If you are interested in providing data, funding, field work, or becoming part of an NHD working group, please contact Robert Hames of the MassDEP GIS Program by email at Robert.Hames@state.ma.us.

Available statewide, the hydrography (arc and polygon feature classes) is stored in ArcSDE as **HYDRO25K_POLY** and **HYDRO25K_ARC**.

PRODUCTION

The DLG quadrangles were converted into Arc/INFO coverages and projected into the Massachusetts State Plane Coordinate System. The long list of items (MAJOR1, MINOR1, MAJOR2, MINOR2...) was then concatenated to a more simplified coding system. For each feature MINORn was truncated to three characters and linked to the other minor codes to create MINOR_TOT. For example, a submerged (612) wetland (111) is now coded MINOR_TOT = 612111. The original MAJORn, MINORn pairs are no longer part of the attribute tables.

Quadrangles covering Nantucket and Martha's Vineyard were completely digitized from the 1:25,000 USGS quadrangles. Though not as thoroughly coded as the 1:25,000 DLGs, the linework is all at 1:25,000.

The scanned quadrangles were automated in-house by scanning USGS mylar separates at 500 dots per inch. The resulting images were vectorized in GRID and then edited in ARCEDIT. Features missing from the blue line separate (i.e. dams or man-made shore) were digitized from the paper quadrangles.

Four quads along the Massachusetts-Connecticut border were obtained from the Connecticut DEP and projected to the Massachusetts State Plane Coordinate System. See the table below for details on the source of each quad.

An ongoing project by the MassDEP GIS Program to redelineate surface water supply watersheds using digital terrain models is adding additional streams within the newly delineated watersheds. These streams are from the MassDEP Wetlands datalayer with some additional on screen digitizing from the 2005 Color OrthoPhotos. Streams added from this process are generally coded as intermittent unless field verification proves otherwise. In 2007 the outlines of all reservoirs were replaced with those from the MassDEP Wetlands datalayer so either dataset can be used with the SWP Zones datalayer. In 2008 field verified streams for the Wachusett Reservoir watershed were provided by DCR West Boylston GIS staff and added.

Existing USGS 1:25000 Hydrography

Hydrologic Connections from DEP Wetlands

MassDEP Hydrography with Local Resolution Streams

EDITING

All of the digitized quadrangles were checkplotted at 1:25,000. The 1:25,000 DLG quadrangles were randomly checkplotted. Each of the quadrangles was edgematched to its neighboring quads. The scanned hydrography was compared both to the source mylar and to the paper quadrangles to ensure completeness.

For the December 2003 update, the layer was extracted from the Quad library and dissolved into a single coverage, which was dissolved on all items after the SOURCE field was dropped. Most slivers, gaps, and other artifacts left over from the quad tiling scheme were removed. Some codes were also refined. Edits were preformed by MassGIS and Department of Environmental Protection (DEP) GIS Group staff.

The high resolution streams from the DEP Wetlands datalayer were field verified and edited mostly in the upper reaches of the watersheds where they crossed roads or if the natural flow wasn't apparent.

ATTRIBUTES

The layer contains both a polygon and arc feature class. The modified DLG coding scheme is extensive and includes a wide variety of features, including ponds, cranberry bogs, impoundments, wetlands, tidal flats, dams, streams, and aqueducts. Only data from the DLGs have been coded this completely. Data from other sources have been coded to include ponds and streams and in the case of data from the scanned quads, wetlands.

Pond and Lake Identification System (PALIS) IDs are unique codes which were added to ponds and lakes by DEP GIS in conjunction with the DEP Division of Watershed Management using identification codes developed by the Pond and Lakes Information System. For historical reasons, some wetland polygons have PALIS IDs. PALIS IDs were also given to impoundment areas along rivers and when necessary closure lines were added.

Since the data now comes from different sources, the attribute SOURCE was added to differentiate which program/entity provided the feature. An attribute for the approximate NHD resolution was added to aid in future input into NHD.

The items in the polygon attribute table are:

MINOR_TOT	Text (15)	Concatenated feature code
POLY_CODE	Number (10)	Generalized code based on MINOR_TOT simplified to these 9 codes: **0** - LAND/ISLAND/DAM/AQUEDUCT **1** - RESERVOIR (with PWSID) **2** - WETLAND, MARSH, SWAMP, BOG **3** - SUBMERGED WETLANDS **4** - CRANBERRY BOG **5** - SALT WETLANDS **6** - LAKE, POND, WIDE RIVER, IMPOUNDMENT **7** - TIDAL FLATS, SHOALS **8** - BAY, OCEAN **9** - INUNDATED AREA
PWSID	Text (11)	DEP public water supply identification number
PALIS_ID	Number (6)	A unique ID from the Ponds and Lakes Information System
SOURCE	Text (12)	Program Source for the feature: **USGS/MGIS** – Original 1:25,000 Hydrography **DEPGIS** – Added from DEP Wetlands by DEP GIS staff **DCRGIS** – From DEP Wetlands and edits/verified by DCR staff
MINOR_NUM	Number (15)	Same as MINOR_TOT, as integer
RESOLUTION	Text (12)	NHD corresponding resolution: **HIGH** - Nominally > 1:25,000 **LOCAL** - < 1:25,000, nominally 1:12,000, some at 1:5,000

The items in the arc attribute table are:

MINOR_TOT	Text (12)	Concatenated feature code
ARC_CODE	Number (10)	Generalized code based on MINOR_TOT simplified to these codes: **1** - SHORELINE **2** - CLOSURE LINE **3** - APPARENT WETLAND LIMIT **4** - STREAM **5** - INTERMITTENT STREAM **6** - DITCH, CANAL **7** - AQUEDUCT **8** - DAM **9** - INTERMITTENT/INDEFINITE SHORELINE **10** - MAN-MADE SHORELINE **11** - CHANNEL IN WATER **99** - TRANSPORT ARC

			0 - UNKNOWN
PWSID	Text (11)	C	DEP public water supply identification number
SOURCE	Text (12)		Program Source for the feature: **USGS/MGIS** – Original 1:25,000 Hydrography **DEPGIS** – Added from DEP Wetlands by DEP GIS staff **DCRGIS** – From DEP Wetlands and edits/verified by DCR staff
MINOR_NUM	Number (12)	I	Same as MINOR_TOT, as integer
RESOLUTION	Text (12)		NHD corresponding resolution: **HIGH** - Nominally > 1:25,000 **LOCAL** - < 1:25,000, nominally 1:12,000, some at 1:5,000

The following table lists all the possible MINOR_TOT/MINOR_NUM codes for a POLYGON FEATURE. These codes have been extracted and concatenated from the USGS DLG major/minor pairs.

MINOR_TOT/NUM	DESCRIPTION
101	RESERVOIR
102	COVERED RESERVOIR
105	INUNDATION AREA
106	FISH HATCHERY/FARM
107	INDUST WATER IMPOUND
109	SEWAGE DISP POND
111	MARSH/WETLAND
114	CRANBERRY BOG
115	FLATS
116	BAY/ESTUARY/GULF
124	FILTRATION POND
400	RAPIDS
401	FALLS
402	GRAVEL PIT W/WATER
404	PUMPING STATION
406	DAM/WEIR
408	SPILLWAY
410	ROCK
412	STREAM
414	DITCH/CANAL
415	AQUEDUCT
416	FLUME
419	CHANNEL IN WATER
421	LAKE OR POND
422	CORAL REEF
999	LAND/ISLAND
7105	INUNDATION AREA
7111	MARSH/WETLAND
101619	RESERVOIR
109421	SEWAGE POND/POND
109611	SEWAGE POND-ABAND
109619	SEWAGE POND
111007	MARSH/WETLAND
111105	MARSH/INUN AREA
111114	MARSH/CRANBERRY BOG
111608	MARSH-SALT
111612	MARSH-SUBMERGED
114007	CRANBERRY BOG
114111	CRANBERRY BOG/MARSH
115007	FLATS
115020	FLATS
115116	FLATS/BAY
115410	FLATS/ROCK
115412	FLATS/STREAM
115421	FLATS/POND
115422	FLATS/CORAL REEF
115616	FLATS-NAVIGABLE
406618	DAM-EARTHEN
410115	ROCKS/FLATS
412111	STREAM/MARSH
412115	STREAM FLATS
412612	STREAM-SUNKEN
415604	AQUEDUCT IN TUNNEL
419115	CHANNEL FLATS
419616	CHANNEL-NAVIGABLE
421007	LAKE OR POND
421101	POND/RESERVOIR
421618	POND-EARTHEN
421619	LAKE OR POND
421625	LAKE OR POND
421628	LAKE OR POND
422115	CORAL REEF/FLATS
610402	INTERMIT PIT W/WATER
610421	INTERMITTENT POND
612111	SUBMERGED MARSH
619101	RESERVOIR
619412	STREAM
619421	LAKE OR POND
101111612	SUBMERGED RES MARSH
101619625	RESERVOIR
105007111	INUN AREA/MARSH
105111007	INUN AREA/MARSH
111007105	MARSH/INUN AREA
111007612	MARSH-SUBMERGED
111007625	MARSH/WETLAND
111105007	MARSH/INUN AREA
111114007	MARSH/CRANBERRY BOG
111612007	MARSH-SUBMERGED
111612421	POND MARSH-SUBMERGED
111612619	MARSH-SUBMERGED
111612625	MARSH-SUBMERGED
114007111	CRANBERRY BOG/MARSH
114111007	CRANBERRY BOG/MARSH
421111007	POND MARSH

101625	RESERVOIR	421111	POND WETLAND	421619625	LAKE OR POND
102111	COV RESERVOIR/MARSH	421609	UNSURVEYED POND	421625619	LAKE OR POND
105007	INUNDATION AREA	421610	POND-INTERMITTENT	619625415	AQUEDUCT
105111	INUN AREA/MARSH	421612	POND-SUNKEN	619625421	LAKE OR POND
				101111612625	RES MARSH-SUBMERGED

The following table lists all the possible MINOR_NUM/MINOR_TOT codes for an ARC FEATURE. These codes have been extracted and concatenated from the USGS DLG major/minor pairs.

MINOR_TOT/NUM	**DESCRIPTION**
109	SEWAGE DISP/FILT BED
200	SHORELINE
201	MANMADE SHORELINE
202	CLOSURE LINE
203	INDEFINITE SHORELINE
204	APPARENT LIMIT
210	TRANSPORT ARC
300	SPRING
401	FALLS
405	WATER INTAKE
406	DAM/WEIR
407	CANAL LOCK
408	SPILLWAY
409	GATE
412	STREAM
414	DITCH/CANAL
415	AQUEDUCT
416	FLUME
419	CHANNEL IN WATER
422	CORAL REEF
605	RIGHT BANK
606	LEFT BANK
999	UNKNOWN OR NEAT LINE
7204	APPARENT LIMIT
9606	LEFT BANK
200009	SHORELINE
200201	MANMADE SHORELINE
200606	LEFT BANK/SHORE

MINOR_TOT/NUM	**DESCRIPTION**
200610	INTERMITTENT SHORE
201002	MANMADE SHORELINE
201009	MANMADE SHORELINE
201020	MANMADE SHORELINE
201605	MANMADE RIGHT BANK
201606	MANMADE LEFT BANK
202412	CLOSURE LINE/STREAM
203625	INDEFINITE SHORE
203627	INDEFINITE SHORE
204007	APPARENT LIMIT
204009	APPARENT LIMIT
406009	DAM/WEIR
406017	DAM/WEIR
406200	DAM/SHORE
406618	DAM/WEIR-EARTHEN
412007	STREAM
412009	STREAM
412020	STREAM
412202	STREAM CLOSURE LINE
412601	STREAM-UNDERGROUND
412604	STREAM-TUNNEL
412609	STREAM-UNSURVEYED
412610	STREAM-INTERMITTENT
414009	DITCH/CANAL
414017	DITCH/CANAL
414610	CANAL-INTERMITTENT
414611	CANAL-ABANDONED
415412	AQUEDUCT/STREAM

MINOR_TOT/NUM	**DESCRIPTION**
415601	AQUEDUCT-UNDERGROUND
415604	AQUEDUCT-TUNNEL
415605	AQUEDUCT RIGHT BANK
415611	AQUEDUCT-ABANDONED
415612	AQUEDUCT-SUBMERGED
601412	UNDERGROUND STREAM
605009	RIGHT BANK
605201	RIGHT BANK-MANMADE
606009	LEFT BANK
606201	LEFT BANK-MANMADE
610200	INTERMITTENT SHORE
610412	INTERMITTENT STREAM
610414	INTERMITTENT CANAL
610421	INTERMITTENT POND
200618406	SHORE/EARTHEN DAM
201009020	MANMADE SHORE
201020002	MANMADE SHORE
201020605	MANMADE RIGHT BANK
201605020	MANMADE RIGHT BANK
201606009	MANMADE LEFT BANK
201606020	MANMADE LEFT BANK
202412617	CLOS LINE/STR UNDERP
406009017	DAM/WEIR
406618017	DAM-EARTHEN
406618200	DAM-EARTHEN/SHORE
412610009	STREAM-INTERMITTENT
412610202	INT STREAM CLOSURE
412610617	INT STRM UNDERPASS

Source of 1:25,000 Hydrography from USGS
Source is for features coded as "USGS/MGIS" in the SOURCE field.
Some of these features may have been modifed by on-screen digitizing.

(USGS QUAD# - SOURCE)
Source Key:
DLG = USGS Digital Line Graph
SCN = Scanned USGS mylar
DIG = Digitized from 1:25,000 USGS topographic maps
CNR = Data from Connecticut modified by MassGIS

1 - DLG	27 - SCN	53 - DLG	79 - DLG	105 - SCN	130 - DLG	155 - SCN	174 - DLG
2 - SCN	28 - DLG	54 - SCN	80 - DLG	106 - SCN	131 - DLG	156 - DLG	174-N - DLG
3 - SCN	29 - SCN	55 - SCN	81 - DLG	107 - SCN	132 - DLG	157 - DIG	175 - DLG
4 - DLG	30 - SCN	56 - SCN	82 - DLG	108 - SCN	132-S - DLG	158 - DIG	176 - DLG
5 - DLG	31 - SCN	57 - SCN	83 - DLG	109 - DLG	133 - DLG	158-S - DIG	176-S - DLG
6 - DLG	32 - SCN	58 - CNR	84 - DLG	110 - DLG	134 - DLG	159-E - DLG	177 - DIG
7 - DLG	33 - SCN	59 - DLG	85 - DLG	111 - DLG	135 - DLG	159-W - DLG	178 - DLG
8 - SCN	34 - DLG	60 - SCN	86 - DLG	112 - DLG	136 - DLG	160 - DLG	179 - DLG
9 - SCN	35 - SCN	61 - SCN	87 - DLG	113 - DLG	137 - DLG	161 - DLG	179-N - DLG
10 - DLG	36 - SCN	62 - SCN	88 - SCN	114 - SCN	138 - DLG	161-E - DLG	180 - DIG
11 - SCN	37 - SCN	63 - SCN	89 - DLG	115 - SCN	139 - DLG	162 - DLG	180-S - DIG
12 - SCN	38 - SCN	64 - DLG	90 - DLG	116 - SCN	140 - DLG	163 - DLG	181 - DLG
13 - SCN	39 - SCN	65 - DLG	91 - DLG	117 - SCN	141 - SCN	164 - DLG	182 - DLG
14 - SCN	40 - CNR	66 - SCN	92 - DLG	118 - SCN	142 - SCN	165 - DLG	182-E - DLG
15 - SCN	41 - SCN	67 - SCN	93 - DLG	119 - DLG	143 - SCN	166 - DIG	183 - DLG
16 - DLG	42 - SCN	68 - SCN	94 - SCN	120 - DLG	144 - DLG	167 - DIG	184 - DLG
17 - SCN	43 - SCN	69 - SCN	95 - DLG	121 - DLG	145 - DIG	168-E - DLG	184-N - DLG
18 - SCN	44 - SCN	70 - DLG	96 - DLG	121-S - DLG	146 - DLG	168-W - DLG	185 - DLG
19 - SCN	45 - SCN	71 - DLG	97 - SCN	122 - DLG	147 - DLG	169 - DLG	186 - DLG
20 - SCN	46 - CNR	72 - DLG	98 - SCN	123 - DLG	148 - DLG	170 - DLG	186-W - DLG
21 - SCN	47 - DLG	73 - DLG	99 - SCN	124 - DLG	149 - DLG	171 - DLG	187 - DIG
22 - DLG	48 - SCN	74 - DLG	100 - SCN	125 - SCN	150 - DLG	172 - DLG	187-E - DIG
23 - SCN	49 - SCN	75 - DLG	101 - SCN	126 - SCN	151 - DLG	173 - DIG	188 - DIG
24 - SCN	50 - SCN	76 - DLG	102 - DLG	127 - SCN	152 - DLG	173-E - DIG	188-E - DIG
25 - SCN	51 - SCN	77 - DLG	103 - DLG	128 - SCN	153 - SCN	173-S - DIG	188-S - DIG
26 - SCN	52 - CNR	78 - DLG	104 - SCN	129 - SCN	154 - SCN	173-SE - DIG	188-SE - DIG
							189 - DLG

Note that theese hydrography layers do not have feature names (e.g., "Quabbin Reservoir" or "Connecticut River") as attributes. To label the features in these layers, use the annotation in the HYDRO subclass of the Geographic Place Names layer.

MAINTENANCE
The MassDEP GIS Program maintains the MassDEP Hydrography layers.

The 2007 update includes new streams within the watersheds of active sources in Lynn, Peabody, Danvers, Middleton, Beverly, Manchester, and Ipswich.

The 2008 update includes new streams within the watersheds of active sources in Gloucester, Rockport, Ipswich, Newburyport, Amesbury, Haverhill, Andover, North Andover, Wakefield, Winchester, Woburn, Burlington, Lincoln, Concord, Leominster, Fitchburg, Worcester, Greenfield and the entire Wachusett Reservoir watershed. Please refer to the SWP Watersheds and SWP Zones datalayers for additional information on this project.

For other hydrographic features, also see the DEP Wetlands (1:12,000), Hydrography (1:100,000) and Major Ponds and Major Streams layers.

Appendix B: The Developed Ontology in OWL

```
<rdf:RDF
    xmlns="http://www.geog.psu.edu/Hydrofeatures.owl#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:owl="http://www.w3.org/2002/07/owl#">
  <owl:Ontology rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl"/>
  <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#marsh">
    <rdfs:subClassOf>
      <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
        </owl:onProperty>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#stream"/>
        </owl:someValuesFrom>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#lake"/>
        </owl:someValuesFrom>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#physiographic_feature">
    <owl:disjointWith>
      <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
    </owl:disjointWith>
  </owl:Class>
  <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#stream">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#swamp"/>
        </owl:someValuesFrom>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#overlaps"/>
        </owl:onProperty>
        <owl:allValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
        </owl:allValuesFrom>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#estuary"/>
        </owl:someValuesFrom>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#gulf"/>
        </owl:someValuesFrom>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
```

```
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#flats"/>
        </owl:someValuesFrom>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#lake"/>
        </owl:someValuesFrom>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
        <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#marsh"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#wetland"/>
        </owl:someValuesFrom>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#bay"/>
        </owl:someValuesFrom>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#overlaps"/>
        </owl:onProperty>
        <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#land_region">
    <owl:disjointWith>
      <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
    </owl:disjointWith>
  </owl:Class>
  <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#lake">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#stream"/>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#bay"/>
        </owl:someValuesFrom>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#wetland"/>
        </owl:someValuesFrom>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
```

```
      <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#inside"/>
        </owl:onProperty>
        <owl:allValuesFrom>
          <owl:Class>
            <owl:unionOf rdf:parseType="Collection">
              <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#land_region"/>
              <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#physiographic_feature"/>
            </owl:unionOf>
          </owl:Class>
        </owl:allValuesFrom>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#swamp"/>
        </owl:someValuesFrom>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
        <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#marsh"/>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#bay">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#channel"/>
        </owl:someValuesFrom>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#pond"/>
        </owl:someValuesFrom>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:cardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:cardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#overlaps"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#flats"/>
        </owl:someValuesFrom>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:allValuesFrom>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
        </owl:allValuesFrom>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#overlaps"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
```

```
      <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#stream"/>
      <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:someValuesFrom>
        <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#reservoir"/>
      </owl:someValuesFrom>
      <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#lake"/>
      <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#estuary">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#stream"/>
      <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#channel">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#bay"/>
      <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#wetland">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#stream"/>
      <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#lake"/>
      <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#pond">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#bay"/>
      <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#reservoir">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#bay"/>
      <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
    </owl:Restriction>
```

```
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#swamp">
    <rdfs:subClassOf>
      <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
        <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#stream"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
        <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#lake"/>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature">
    <owl:disjointWith rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#physiographic_feature"/>
    <owl:disjointWith rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#land_region"/>
  </owl:Class>
  <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#flats">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#stream"/>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#bay"/>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
  </owl:Class>
  <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#gulf">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#stream"/>
        <owl:onProperty rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#evolves_into"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
  </owl:Class>
  <owl:ObjectProperty rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#overlaps">
    <rdfs:domain rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
    <rdfs:range rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#inside">
    <rdfs:domain rdf:resource="http://www.geog.psu.edu/Hydrofeatures.owl#hydrographic_feature"/>
    <rdfs:range>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#land_region"/>
          <owl:Class rdf:about="http://www.geog.psu.edu/Hydrofeatures.owl#physiographic_feature"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:range>
  </owl:ObjectProperty>
</rdf:RDF>
```

Appendix C: Metadata for the Florida Dataset

This metadata was originally obtained from the Florida Geographic Data Library (http://www.fgdl.org/).

Geodataset Name: GNIS_JUL06
Geodataset Type: SHAPEFILE
Geodataset Feature: POINT
Feature Count: 36097

GENERAL DESCRIPTION:

This dataset contains information about physical and cultural geographic features of all types, but not including roads and highways. The database holds the Federally recognized name of each feature and defines the feature location by state, county, USGS topographic map, and geographic coordinates.

DATA SOURCE(S): United States Geological Survey
SCALE OF ORIGINAL SOURCE MAPS: N/A
DATE OF AUTOMATION OF SOURCE: 20060719
GEODATASET EXTENT: STATE OF FLORIDA

FEATURE ATTRIBUTE TABLES:

Datafile Name: GNIS_JUL.DBF

ITEM NAME	WIDTH	TYPE	N. DECIMAL DEGREES
OBJECTID	4	OID	---
FEATURE_ID	10	Number	---
FEATURE_NA	71	String	---
CLASS	16	String	---
ST_ALPHA	8	String	---
ST_NUM	8	Number	---
COUNTY	11	String	---
COUNTY_NUM	11	Number	---
PRI_LATDMS	16	String	---
PRI_LONDMS	16	String	---
PRI_LATDEC	19	Number	7
PRI_LONDEC	19	Number	7
SOU_LATDMS	15	String	---
SOU_LONDMS	15	String	---
SOU_LATDEC	14	String	---
SOU_LONDEC	14	String	---
ELEV_METER	11	Number	---

MAP_NAME	27	String	---
DESCRIPT	71	String	---
FGDLAQDATE	8	Date	---
SHAPE	4	Geometry	---
AUTOID	4	Integer	---

FEATURE ATTRIBUTE TABLES CODES AND VALUES:

Item	Item Description
OBJECTID	Internal feature number.
FEATURE_ID	Feature Identifier. A permanent, unique identifying number for a geographic feature record.
FEATURE_NA	Feature Name. Recognized as the official name of the US Board on Geographic Names for use throughout the Federal Government. The official name is established either by policy of by decision of the Board on Geographic Names.
CLASS	Feature Class. A feature class is a designation for a group of features in a broadly defined descriptive category. Airport = Manmade facility maintained for the use of aircraft (airfield, airstrip, landing field strip). Arch = Natural arch-like opening in a rock mass (bridge, natura bridge, sea arch). Area = Any one of several really extesive natural features not included in other categories (badlands, barren, delta, fan, garden). Bar = Natural accumulation of sand, gravel, or alluvium forming an underwater or exposed embankment (ledge, reef, sandbar, shoal, spit). Basin = Natural depression or relatively low area enclosed by higher land (amphitheater, cirque, pit, sink). Bay = Indentation of a coastline or shoreline enclosing a part of a body of water; a body of water partly surrounded by land (arm, bight, cove, estuary, gulf, inlet, sound). Beach = The sloping shore along a body of water that is washed by waves or tides and is usually covered by sand or gravel (coast, shore, strand). Bend = Curve in the course of a stream and (or) the land within the curve; a curve in a linear body of water (bottom, loop, meander). Bridge = Manmade structure carrying a trail, road, or other transportation system across a body of water or depression (causeway, overpass, trestle). Building = A manmade structure with walls and a roof for protection of people and (or) materials, but not including church, hospital, or school. Canal = Manmade waterway used by watercraft or for drainage, irrigation, mining, or water power (ditch, lateral). Cape = Projection of land extending into a body of water (lea, neck, peninsula, point). Cave = Natural underground passageway or chamber, or a hollowed out cavity in the side of a cliff (cavern, grotto).

Cemetary = A place or area for burying the dead (burial, burying ground, grave, memorial garden).
Channel = Linear deep part of a body of water through which the main volume of water flows and is frequently used as a route for watercraft (passage, reach, strait, thoroughfare, throughfare).
Church = Building used for religious worship (chapel, mosque, synagogue, tabernacle, temple).
Civil = A political division formed for administration purposes (borough, county, municipio, parish, town, township).
Cliff = Very steep or vertical slope (bluff, crag, head, headland, nose, palisades, precipice, promontory, rim, rimrock).
Crossing = A place where two or more routes of transportation form a junction or intersection (overpass, underpass).
Dam = Water barrier or embankment built across the course of a stream or into a body of water to control and (or) impound the flow of water (breakwater, dike, jetty).
Falls = Perpendicular or very steep fall of water in the course of a stream (cascade, cataract, waterfall).
Flat = Relative level area within a region of greater relief (clearing, glade, playa).
Forest = Bounded area of woods, forest, or grassland under the administration of a political agency (see "woods") (national forest, national grasslands, State forest).
Gap = Low point or opening between hills or mountains or in a ridge or mountain range (col, notch, pass, saddle, water gap, wind gap).
Gut = Relatively small coastal waterway connecting larger bodies of water or other waterways (creek, inlet, slough).
Harbor = Sheltered area of water where ships or other watercraft can anchor or dock (hono, port, roads, roadstead).
Hospital = Building where the sick or injured may receive medical or surgical attention (infirmary).
Island = Area of dry or relatively dry land surrounded by water or low wetland (archipelago, atoll, cay, hammock, hummock, isla, isle, key, moku, rock).
Lake = Natural body of inland water (backwater, lac, lagoon, laguna, pond, pool, resaca, waterhole).
Levee = Natural or manmade embankment flanking a stream (bank, berm).
Locale = Place at which there is or was human activity; it does not include populated places, mines, and dams (battlefield, crossroad, camp, farm, ghost town, landing, railroad siding, ranch, ruins, site, station, windmill).
Military (Histor = Place or facility formerly used for various aspects of or relating to military activity.
Military = Place or facility used for various aspects of or relating to military activity.
Mine = Place or area from which commercial minerals are or were removed from the Earth; not including oilfield (pit, quarry, shaft).

Park = Place or area set aside for recreation or preservation of a cultural or natural resource and under some form of government administration; not including National or State forests or Reserves (national historical landmark, national park, State park, wilderness area).
Populated Place = (Formerly abbreviated as ppl) place or area with clustered or scattered buildings and a permanent human population (city, settlement, town, village).
Post Office = (Formerly abbreviated as PO) an official facility of the U.S. Postal Service used for processing and distributing mail and other postal material.
Reserve = A tract of land set aside for a specific use (does not include forests, civil divisions, parks).
Reservoir = Artificially impounded body of water (lake, tank).
Ridge = Elevation with a narrow, elongated crest which can be part of a hill or mountain (crest, cuesta, escarpment, hogback, lae, rim, spur).
School = Building or group of buildings used as an institution for study, teaching, and learning (academy, college, high school, university).
Spring = Place where underground water flows naturally to the surface of the Earth (seep).
Stream = Linear body of water flowing on the Earth's surface (anabranch, awawa, bayou, branch, brook, creek, distributary, fork, kill, pup, rio, river, run, slough).
Summit = Prominent elevation rising above the surrounding level of the Earth's surface; does not include pillars, ridges, or ranges (ahu, berg, bald, butte, cerro, colina, cone, cumbre, dome, head, hill, horn, knob, knoll, mauna, mesa, mesita, mound, mount, mountain, peak, puu, rock, sugarload, table, volcano).
Swamp = Poorly drained wetland, fresh or saltwater, wooded or grassy, possibly covered with open water (bog, cienega, marais, marsh, pocosin).
Tower = A manmade structure, higher than its diameter, generally used for obersavation, storage, or electronic transmission.
Trail = Route for passage from one point to another; does not inlcude roads or highways (jeep trail, path, ski trail).
Unknown =
Valley = Linear depression in the Earth's surface that generally slopes from one end to the other (barranca, canyon, chasm, cove, draw, glen, gorge, gulch, gulf, hollow, ravine).
Well = Manmade shaft or hole in the Earth's surface used to obtain fluid or gaseous materials.
Woods = Small area covered with a dense growth of trees; does not include an area of trees under the administration of a political agency (see "forest").

ST_ALPHA	The state containing the geographic feature.
ST_NUM	The FIPS number representing the state the feature is located within.
COUNTY	The county containing the geographic feature.
COUNTY_NUM	The FIPS number representing the county the feature is located within.

PRI_LATDMS	Official feature latitude location, NAD 83 DMS-degrees/minutes/seconds
PRI_LONDMS	Official feature longitude location, NAD 83 DMS-degrees/minutes/seconds
PRI_LATDEC	Official feature latitude location, NAD 83 DEC-decimal degrees.
PRI_LONDEC	Official feature longitude location, NAD 83 DEC-decimal degrees.
SOU_LATDMS	Source latitude coordinates of linear feature only (Class = Stream, Valley, Arroyo), NAD 83; DMS-degrees/minutes/seconds.
SOU_LONDMS	Source longitude coordinates of linear feature only (Class = Stream, Valley, Arroyo), NAD 83; DMS-degrees/minutes/seconds.
SOU_LATDEC	Source latitude coordinates of linear feature only (Class = Stream, Valley, Arroyo), NAD 83; DEC-decimal degrees.
SOU_LONDEC	Source longitude coordinates of linear feature only (Class = Stream, Valley, Arroyo), NAD 83; DEC-decimal degrees.
ELEV_METER	The elevation above sea level of the feature at the primary point (positive number) or depth of a feature at the lowest point below sea level (negative number).
MAP_NAME	The name of the standard USGS 7.5 x 7.5 degree quadrangle map(s) containing the representation of the feature.
DESCRIPT	GeoPlan added field based on FEATURE_NA Basin = Natural depression or relatively low area enclosed by higher land(amphitheater, cirque, pit, sink). Bay = Indentation of a coastline or shoreline enclosing a part of a body of water; a body of water partly surrounded by land (arm, bight, cove, estuary, gulf, inlet, sound). Beach = The sloping shore along body of water that is washed by waves or tides and is usually covered by sand or gravel (coast, shore, strand).
FGDLAQDATE	Date FGDL acquired the data.
SHAPE	Feature geometry.
AUTOID	Unique ID added by GeoPlan

USER NOTES:

This data is provided 'as is'. GeoPlan relied on the integrity of the original data layer's topology

This data is provided 'as is' by GeoPlan and is complete to our knowledge.

GeoPlan relied on the integrity of the attribute information within the original data.

The Geographic Names Information System (GNIS) is the Federal standard for geographic nomenclature. The U.S. Geological Survey developed the GNIS for the U.S. Board on Geographic Names as the official repository of domestic geographic names data; the official vehicle for geographic names use by all departments of the Federal Government; and the source for applying geographic names to Federal electronic and printed products.

The GNIS contains information about physical and cultural geographic features of all types in the United States, associated areas, and Antarctica, current and historical, but not including roads and highways. The database holds the Federally recognized name of each feature and defines the feature location by state, county, USGS topographic map, and geographic coordinates. Other attributes include names or spellings other than the official name, feature designations, feature classification, historical and descriptive information, and for some categories the geometric boundaries.

This data is provided 'as is' and its horizontal positional accuracy has not been verified by GeoPlan

A note about data scale:

Scale is an important factor in data usage. Certain scale datasets are not suitable for some project, analysis, or modeling purposes. Please be sure you are using the best available data.

1:24000 scale datasets are recommended for projects that are at the county level.
1:24000 data should NOT be used for high accuracy base mapping such as property parcel boundaries.
1:100000 scale datasets are recommended for projects that are at the multi-county or regional level.
1:125000 scale datasets are recommended for projects that are at the regional or state level or larger.

Vector datasets with no defined scale or accuracy should be considered suspect. Make sure you are familiar with your data before using it for projects or analysis. Every effort has been made to supply the user with data documentation. For additional information, see the References section and the Data Source Contact section of this documentation. For more information regarding scale and accuracy, see our webpage at: http://geoplan.ufl.edu/education.html

REFERENCES:

U.S. Board on Geographic Names Definition list
http://geonames.usgs.gov/domestic/feature_class.htm

DATA LINEAGE SUMMARY:

Data was downloaded from the United States Geological Survey: U.S. Board on Geographic Names website in text format to GeoPlan in September 2006. The original data extent was limited to the State of
Florida. GeoPlan, during the QA/QC process included the following aspects:

(1) The original dataset name was FL_DECI.txt
(2) The text file was converted to .DBF format. The .DBF file was then converted to a shapefile using the Add XY tool in ArcMap.
(3) The original points were recorded in NAD83. GeoPlan projected the entire dataset to FGDL Albers.
(4) The following points plotted outside the State of Florida or were located incorrectly. The corresponding Feature ID's of the deleted points are: 296517, 296521, 296523, 296814, 297009, 297717, 298049, 298165, 298211, 298325, 298736, 298979, 299184, 300608, 301010, 301041, 301312, 301880, 301889, 302374, 302376, 302384, 302456, 302754, 302829, 302904, 302930, 302963, 303008, 303059, 303372, 303382, 303751, 303753, 304298, 304641, 307146, 308181, 308609, 309281, 310263, 308529, 298711, 302991, and 300536.
(5) Deleted points which plotted outside of the physical State of Florida.
(6) Added the FGDLAQDATE field.
(7) Added DESCRIPT item based on FEATURE_NA. Process Date: 20061016

MAP PROJECTION PARAMETERS:

Projection	ALBERS
Datum	HPGN
Units	METERS
Spheroid	GRS1980
1st Standard Parallel	24 0 0.000
2nd Standard Parallel	31 30 0.000
Central Meridian	-84 00 0.000
Latitude of Projection's Origin	24 0 0.000
False Easting (meters)	400000.00000
False Northing (meters)	0.00000

DATA SOURCE CONTACT (S):

	U.S. Board on Geographic Names, USGS
Name:	USGS
Abbr. Name:	12201 Sunrise Valley Drive, MS 523
Address:	Reston, Virginia
	20192-0523
	703-648-4544
Phone:	

Web site: http://geonames.usgs.gov/domestic/download_data.htm
E-mail: gnis_manager@usgs.gov
Contact Person:
Phone:
E-mail:

FGDL CONTACT:

Name: FLORIDA GEOGRAPHIC DATA LIBRARY
Abbr. Name: FGDL
Address: Florida Geographic Data Library
431 Architecture Building
PO Box 115706
Gainesville, FL 32611-5706
Web site: http://www.fgdl.org

Contact FGDL:
Technical Support: http://www.fgdl.org/fgdlfeed.html
FGDL Frequently Asked Questions: http://www.fgdl.org/fgdlfaq.html
FGDL Mailing Lists: http://www.fgdl.org/fgdl-l.html
For FGDL Software: http://www.fgdl.org/software.html

Appendix D: Metadata for the Alaska Dataset

This metadata was originally obtained from the Alaska State Geo-spatial Data Clearinghouse (ASGDC) (http://www.asgdc.state.ak.us/)

Identification_Information:
 Citation:
  Citation_Information:
   Originator: Alaska Department of Natural Resources - Land Records Information Section
   Publication_Date: 1967
   Title: USGS Alaska Place Names
   Geospatial_Data_Presentation_Form: vector digital data
   Online_Linkage: http://dnr.alaska.gov/SpatialUtility/SUC?cmd=extract&layerid=9
 Description:
  Abstract:
   The usgsname coverage was generated from the latitude, longitude coordinates provided by USGS. These locations reflect the coordinates and names cited in the Dictionary of Alaska Place Names. The date of the Dictionary is 1967, while it was automated in more recent years.

   These are points only, with corresponding attributes. Note, the names of these locations have been automated into Arc/Info annotation in a separate coverage.

  Purpose: Identify location of place names.
 Time_Period_of_Content:
  Time_Period_Information:
   Single_Date/Time:
    Calendar_Date: 1967
  Currentness_Reference: Publication date.
 Status:
  Progress: Complete
  Maintenance_and_Update_Frequency: As needed.
 Spatial_Domain:
  Bounding_Coordinates:
   West_Bounding_Coordinate: -180.000000
   East_Bounding_Coordinate: 180.000000
   North_Bounding_Coordinate: 76.786209
   South_Bounding_Coordinate: 42.509212
 Keywords:
  Theme:
   Theme_Keyword_Thesaurus: ISO 19115 Topic Category
   Theme_Keyword: sand bars
   Theme_Keyword: beaches
   Theme_Keyword: basins
   Theme_Keyword: bays
   Theme_Keyword: bends
   Theme_Keyword: buildings
   Theme_Keyword: canals
   Theme_Keyword: capes
   Theme_Keyword: cemetery
   Theme_Keyword: channels
   Theme_Keyword: churches
   Theme_Keyword: civil
   Theme_Keyword: cliffs
   Theme_Keyword: craters
   Theme_Keyword: dams
   Theme_Keyword: falls

Theme_Keyword: flats
Theme_Keyword: forests
Theme_Keyword: glaciers
Theme_Keyword: harbors
Theme_Keyword: hospitals
Theme_Keyword: islands
Theme_Keyword: lakes
Theme_Keyword: locations
Theme_Keyword: military
Theme_Keyword: mines
Theme_Keyword: parks
Theme_Keyword: pillars
Theme_Keyword: populated places
Theme_Keyword: ranges
Theme_Keyword: rapids
Theme_Keyword: reserves
Theme_Keyword: reservoirs
Theme_Keyword: ridges
Theme_Keyword: schools
Theme_Keyword: seas
Theme_Keyword: springs
Theme_Keyword: streams
Theme_Keyword: summits
Theme_Keyword: swamps
Theme_Keyword: trails
Theme_Keyword: valleys
Theme_Keyword: wells
Theme_Keyword: woods
Theme_Keyword: peaks
Theme_Keyword: base map
Theme_Keyword: physical features
Theme_Keyword: geographic locations
Theme_Keyword: gazetteer

Place:
Place_Keyword_Thesaurus: none
Place_Keyword: Alaska
Access_Constraints:
To ensure distribution of the most current public information, please refer requests for data or products to the Alaska Department of Natural Resources, Land Records Information Section.

Use_Constraints:
The recommended scale for this depends on the feature type. Point features may be very accurate, while linear or polygon features, being represented in this coverage by a point, may be less accurate.

Any hardcopies utilizing these data sets shall clearly indicate their source. If the user has modified the data in any way they are obligated to describe the types of modifications they have performed on the hardcopy map. User specifically agrees not to misrepresent these data sets, nor to imply that changes they made were approved by Alaska Department of Natural Resources.
Point_of_Contact:
Contact_Information:
Contact_Person_Primary:
Contact_Person: GIS Public Access Coordinator
Contact_Organization: AK Department of Natural Resources - Land Records Information Section
Contact_Position: GIS Public Access Coordinator
Contact_Address:
Address_Type: mailing and physical address
Address: 550 W. 7th Suite 706
City: Anchorage
State_or_Province: AK
Postal_Code: 99501
Country: USA

Contact_Voice_Telephone: 907 269 8833
Contact_Facsimile_Telephone: 907 269 8920
Contact_Electronic_Mail_Address: GIS_Public_Access@dnr.state.ak.us
Hours_of_Service: 800-1600 AST
Contact_Instructions: See Distribution_Information
Data_Quality_Information:
Attribute_Accuracy:
Attribute_Accuracy_Report:
Were a few records where the text was out of order. For example, in most cases the lakes were listed as "Lake Clark". In some cases it was listed as "Clark, Lake". A few features were missing from the designated feature, could be attributed to the age of the source document.
Logical_Consistency_Report: Point features present.
Completeness_Report: Unknown for the data entry. See Attribute_Accuracy_Report.
Positional_Accuracy:
Horizontal_Positional_Accuracy:
Horizontal_Positional_Accuracy_Report: See Use Constraints.
Lineage:
Source_Information:
Source_Citation:
Citation_Information:
Originator: USGS
Publication_Date: 1967
Title: Dictionary of Alaska Place Names
Series_Information:
Series_Name: Geological Survey Professional Paper 567
Issue_Identification: Unknown
Publication_Information:
Publication_Place: Washington, DC
Publisher: US Government Printing Office
Source_Scale_Denominator: 250000
Type_of_Source_Media: Book
Source_Time_Period_of_Content:
Time_Period_Information:
Single_Date/Time:
Calendar_Date: 1967
Source_Currentness_Reference: Publication date.
Source_Citation_Abbreviation: None
Source_Contribution: Source document for electronic listing.
Process_Step:
Process_Description:
USGS provided DNR with a table of lat/long, names, and designations (such as bay, beach, etc.) DNR converted these to point locations. This information was subsequently used as the base for the annotation coverages available from DNR.
Process_Date: 19950412
Process_Step:
Process_Description: USGSNAME coverage imported into geodatabase.
Process_Date: unknown
Process_Step:
Process_Description: Geodatabase feature classes data copied to materialized view.
Process_Date: 20051010
Process_Step:
Process_Description: Materialized view data extracted and converted to Shape format.
Process_Date: unknown
Spatial_Data_Organization_Information:
Direct_Spatial_Reference_Method: Vector
Point_and_Vector_Object_Information:
SDTS_Terms_Description:
SDTS_Point_and_Vector_Object_Type: Entity point
Point_and_Vector_Object_Count: 25876
Spatial_Reference_Information:
Horizontal_Coordinate_System_Definition:
Planar:

Map_Projection:
Map_Projection_Name: Albers Conical Equal Area
Albers_Conical_Equal_Area:
Standard_Parallel: 55.000000
Standard_Parallel: 65.000000
Longitude_of_Central_Meridian: -154.000000
Latitude_of_Projection_Origin: 50.000000
False_Easting: 0.000000
False_Northing: 0.000000
Planar_Coordinate_Information:
Planar_Coordinate_Encoding_Method: coordinate pair
Coordinate_Representation:
Abscissa_Resolution: 0.000100
Ordinate_Resolution: 0.000100
Planar_Distance_Units: meters
Geodetic_Model:
Horizontal_Datum_Name: North American Datum of 1983
Ellipsoid_Name: Geodetic Reference System 80
Semi-major_Axis: 6378137.000000
Denominator_of_Flattening_Ratio: 298.257222
Entity_and_Attribute_Information:
Detailed_Description:
Entity_Type:
Entity_Type_Label: USGS_PLACE_NAME
Entity_Type_Definition: Feature class
Entity_Type_Definition_Source: ESRI
Attribute:
Attribute_Label: FID
Attribute_Definition: Internal feature number.
Attribute_Definition_Source: ESRI
Attribute_Domain_Values:
Unrepresentable_Domain: Sequential unique whole numbers that are automatically generated.
Attribute:
Attribute_Label: Shape
Attribute_Definition: Feature geometry.
Attribute_Definition_Source: ESRI
Attribute_Domain_Values:
Unrepresentable_Domain: Coordinates defining the features.
Attribute:
Attribute_Label: GEONAME
Attribute_Definition: Name of feature.
Attribute_Definition_Source: USGS
Attribute_Domain_Values:
Unrepresentable_Domain: Each name is unique.
Attribute:
Attribute_Label: DESIG
Attribute_Definition: Type of feature.
Attribute_Definition_Source: USGS
Attribute_Domain_Values:
Unrepresentable_Domain:

-	1
ARCH	2
AREA	274
BAR	335
BASIN	42
BAY	2146
BEACH	28
BEND	28
BUILDING	7
CANAL	20
CAPE	2238
CEMETERY	16

CHANNEL 417
CHURCH 5
CIVIL 19
CLIFF 148
CRATER 10
DAM 2
FALLS 40
FLAT 31
FOREST 1
GAP 190
GLACIER 595
GUT 161
HARBOR 13
HOSPITAL 2
ISLAND 2102
LAKE 2868
LOCALE 980
MILITARY 16
MINE 163
OTHER 497
PARK 92
PILLAR 26
PPL 408
RANGE 60
RAPIDS 14
RESERVE 5
RESERVOIR 8
RIDGE 164
SCHOOL 75
SEA 1
SPRING 23
STREAM 8551
SUMMIT 2523
SWAMP 7
TRAIL 59
VALLEY 420
WELL 41
WOODS 1

Attribute:
Attribute_Label: COUNTY
Attribute_Definition: Misnomer for boroughs and other boundaries.
Attribute_Definition_Source: USGS
Attribute_Domain_Values:
Unrepresentable_Domain:
- 4
Alaska Peninsula Borough 2
Aleutian Islands 1756
Anchorage 237
Anchorage Borough 136
Anchorge 1
Bethel 908
Bethel Borough 2
Bristol Bay 41
Bristol Bay Borough 4
Dillingham 1212
Fairbanks North Star 538
Fairbanks North Star Borough 3
Faribanks North Star Borough 1
Haines 186
Haines Borough 2
Juneau 519
Juneau Borough 14

Kenai Peninsula	1508
Kenai Peninsula Borough	85
Kenai Peninsula Burough	1
Kenai Pennisula	16
Ketchikan Gateway	341
Ketchikan Gateway Borough	1
Kobuk	3
Kodiak Island	802
Kodiak Island Borough	37
Matanuska-Susitna	1220
Matanuska-Susitna Borough	145
NONE	2
Nome	1708
North Slope	1378
North Slope Borough	39
Northwest Arctic	724
Northwest Artic	38
Prince of WAles-Outer Ketchikan	1
Prince of Wales-Outer Ketchikan	1827
Seldovia (B-4)	1
Sitka	953
Sitka Borough	26
Skagway-Yakutat Borough	2
Skagway-Yakutat-Angoon	1293
Southeast Fairbanks	873
Survey Pass 1:250	1
Valdez-Cordova	2219
Valdez-Cordova Borough	13
Wade Hampton	444
Wrangell-Petersburg	848
Yukon-Koyukuk	3752
Yukon-Koyukuk Borough	5
unknown	3

Attribute:
Attribute_Label: ELEV
Attribute_Definition: Elevation for each feature.
Attribute_Definition_Source: USGS
Attribute_Domain_Values:
Unrepresentable_Domain: Each value is unique.
Distribution_Information:
Distributor:
Contact_Information:
Contact_Organization_Primary:
Contact_Organization: Alaska Department of Natural Resources - Land Records Information Section
Contact_Position: GIS Public Access Coordinator
Contact_Address:
Address_Type: mailing and physical address
Address: 550 W. 7th Suite 706
City: Anchorage
State_or_Province: AK
Postal_Code: 99501
Country: USA
Contact_Voice_Telephone: 907 269 8833
Contact_Facsimile_Telephone: 907 269 8920
Resource_Description: Downloadable Data
Distribution_Liability:
The State of Alaska makes no express or implied warranties (including warranties of merchantability and fitness) with respect to the character, function, or capabilities of the electronic services or products or their appropriateness for any users purposes. In no event will the State of Alaska be liable for any incidental, indirect, special, consequential or other damages suffered by the user or any other person or entity whether from the use of the electronic services or products, any failure thereof or otherwise, and in no event will the State of Alaska s liability to the

requestor or anyone else exceed the fee paid for the electronic service or product.
Standard_Order_Process:
Digital_Form:
Digital_Transfer_Information:
Format_Name: zipped shape file
Digital_Transfer_Option:
Online_Option:
Computer_Contact_Information:
Network_Address:
Network_Resource_Name: http://dnr.alaska.gov/SpatialUtility/SUC?cmd=extract&layerid=9
Fees: none
Metadata_Reference_Information:
Metadata_Date: 20080204
Metadata_Contact:
Contact_Information:
Contact_Organization_Primary:
Contact_Organization: AK Department of Natural Resources - Land Records Information Section
Contact_Position: GIS Public Access Coordinator
Contact_Address:
Address_Type: mailing and physical address
Address: 550 W. 7th Suite 706
City: Anchorage
State_or_Province: AK
Postal_Code: 99501
Country: USA
Contact_Voice_Telephone: 907 269 8833
Contact_Facsimile_Telephone: 907 269 8920
Contact_Electronic_Mail_Address: gis_public_access@dnr.state.ak.us
Hours_of_Service: 800-1600 AST
Metadata_Standard_Name: FGDC Content Standards for Digital Geospatial Metadata
Metadata_Standard_Version: FGDC-STD-001-1998
Metadata_Time_Convention: local time
Metadata_Use_Constraints:
If the user has modified the data in any way they are obligated
to describe the types of modifications they have performed in the
supporting metadata file. User specifically agrees not to imply
that changes they made were approved by the Alaska Department of
Natural Resources.
Metadata_Extensions:
Online_Linkage: http://www.esri.com/metadata/esriprof80.html
Profile_Name: ESRI Metadata Profile