5-2015

Enhanced Place Name Search Using Semantic Gazetteers

Nagalakshmy Vijayasankaran

ENHANCED PLACE NAME SEARCH USING SEMANTIC GAZETTEERS

By

Nagalakshmy Vijayasankaran

B.E., B.Tech., Anna University, 2004

A THESIS

Submitted in Partial Fulfilment of the

Requirements for the Degree of

Master of Science

(in Spatial Information Science and Engineering)

The Graduate School

The University of Maine

May 2015

Advisory Committee:

Dr. Kate Beard-Tisdale, Professor of Spatial Informatics, Advisor

Dr. Max Egenhofer, Professor of Spatial Informatics

Dr. Torsten Hahmann, Assistant Professor of Spatial Informatics

THESIS ACCEPTANCE STATEMENT

On behalf of the Graduate Committee for Nagalakshmy Vijayasankaran, I affirm that this manuscript is the final and accepted thesis. Signatures of all committee members are on file with the Graduate School at the University of Maine, 42 Stodder Hall, Orono, Maine.

Dr.Kate Beard-Tisdale, Professor of Spatial Informatics Date:

LIBRARY RIGHTS STATEMENT

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at The University of Maine, I agree that the Library shall make it freely available for inspection. I further agree that permission for "fair use" copying of this thesis for scholarly purposes may be granted by the Librarian. It is understood that any copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Signature:

Date:

ENHANCED PLACE NAME SEARCH USING SEMANTIC GAZETTEERS

By Nagalakshmy Vijayasankaran

Thesis Advisor: Dr. Kate Beard-Tisdale

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Master of Science
(in Spatial Information Science and Engineering)
May 2015

With the increased availability of geospatial data and efficient geo-referencing services, people are now more likely to engage in geospatial searches for information on the Web. Searching by address is supported by geocoding which converts an address to a geographic coordinate. Addresses are one form of geospatial referencing that are relatively well understood and easy for people to use, but place names are generally the most intuitive natural language expressions that people use for locations. This thesis presents an approach, for enhancing place name searches with a geo-ontology and a semantically enabled gazetteer. This approach investigates the extension of general spatial relationships to domain specific semantically rich concepts and spatial relationships. Hydrography is selected as the domain, and the thesis investigates the specification of semantic relationships between hydrographic features as functions of spatial relationships between their footprints.

A Gazetteer Ontology (GazOntology) based on ISO Standards is developed to associate a feature with a Spatial Reference. The Spatial Reference can be a GeoIdentifier which is a text based representation of a feature usually a place name or zip code or the spatial reference can be a Geometry representation which is a spatial footprint of the feature. A Hydrological Features Ontology (HydroOntology) is developed to model canonical forms of hydrological features and their hydrological relationships. The classes modelled are endurant classes modelled in foundational ontologies such as DOLCE. Semantics of these relationships in a hydrological context are specified in a HydroOntology.

The HydroOntology and GazOntology can be viewed as the semantic schema for the HydroGazetteer. The HydroGazetteer was developed as an RDF triplestore and populated with instances of named hydrographic features from the National Hydrography Dataset (NHD) for several watersheds in the state of Maine. In order to determine what instances of surface hydrology features participate in the specified semantic relationships, information was obtained through spatial analysis of the National Hydrography Dataset (NHD), the NHDPlus data set and the Geographic Names Information System (GNIS). The 9 intersection model between point, line, directed line, and region geometries which identifies sets of relationship between geometries independent of what these geometries represent in the world provided the basis for identifying semantic relationships between the canonical hydrographic feature types.

The developed ontologies enable the HydroGazetteer to answer different categories of queries, namely place name queries involving the taxonomy of feature types, queries on relations between named places, and place name queries with reasoning. A simple user interface to select a hydrological relationship and a hydrological feature name was developed and the results are displayed on a USGS topographic base map. The approach demonstrates that spatial semantics can provide effective query disambiguation and more targeted spatial queries between named places based on relationships such as upstream, downstream, or flows through.

ACKNOWLEDGEMENTS

I would like to thank all the people who have supported me throughout the course of this research. First and foremost, I would like to thank Dr. Kate Beard-Tisdale, my advisor who has been very supportive of me to complete this work. I would like to thank her for her technical guidance, extensive discussions and insights which have played an important role in this work. I would also like to thank her for her confidence in me and giving me this opportunity to complete my education at this institution.

I would like to thank Dr. Max Egenhofer for being a part of my committee and for his valuable comments and suggestions.

I would like to thanks Dr. Torsten Hahmann for having agreed to serve in my thesis committee and review my work.

I would like to thank Dr. Anthony Stefanidis, who gave me the opportunity to pursue my studies at the University of Maine.

I would like to thank my late grandparents who raised me with a lot of love and affection to be the person I am today. I would like to thank my husband who has shared every part of this thesis and has been a solid support for me throughout. I would like to thank my parents for this wonderful life , my brother who has led by example, my sister in law and my family and friends, who have influenced my life with their love, encouragement, experience, knowledge and wisdom and above all for being there for me whenever I turned to them in times of need.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

CHAPTER 1

INTRODUCTION

With the increased availability of geospatial data and efficient geo-referencing services, people are now more likely to engage in geospatial searches for information on the Web. It is quite simple to find common places in day to day life. The advent of online location-based services, such as Google Maps, MapQuest, and other similar services, has made it possible to find geographic locations of addresses and businesses along with routing instructions for navigating between locations. Searching by address is supported by geocoding, which converts an address to a geographic coordinate. Addresses are one form of geospatial referencing that are relatively well understood and easy for people to use, but place names are generally the most intuitive natural language expressions that people use for locations. “Pizza Hut near Houston, TX” is a typical example of a place name search that in this case is qualified by the spatial relation “near”. This query results in all locations of Pizza Hut businesses in the Houston area displayed on a map. In this example, processing the query requires the place name to be converted to a geographic coordinate position. In address geocoding, an address is converted to a geographic coordinate through matching a user’s address to an address location in a street reference file.

Gazetteers are knowledge organization systems that consist of triples (N, F, T), where N corresponds to the place name, F to the geographic footprint and T to the place type (Hill, 2009). Place names are converted to geographic coordinates through the mechanism of gazetteers (Hill & Zheng, 1998). Gazetteers are information sources that enable users to map a place name to a geographic location along with its feature type. Feature type information usually organized as a thesaurus, helps in disambiguating two similar place names. In web-based searches, people may use place names as a keyword, in which case the search relies predominantly on simple text string matching without benefit of the added information provided by a gazetteer.

While the World Wide Web has been very successful in bringing data to the end user through value-enabled services, it still falls short in supporting geospatially enabled searches and reasoning. For better geospatial search and reasoning, we need better supported information retrieval methods that delve deeper than keyword searches, directory services, and page ranks. A new way of organizing and retrieving distributed and inter related data is critical to move the Web from a data repository to an information system (Egenhofer M. , 2002). The primary focus of the Semantic Web (Berners-Lee, Hendler, & Lassila, 2001) is to add semantic meaning to web content that supports interoperability of different data sources and reasoning in the Web. The Semantic Web adds meaning to the data such that it is machine-understandable as well as accessible to human comprehension. A series of standards, such as Extensible Markup Language (XML), Resource Description Framework (RDF), Resource Description Framework Schema (RDFS), Web Ontology language (OWL), and SPARQL query language, have been developed by the World Wide Web Consortium as supporting technologies for the Semantic Web.

Among these semantic web technologies, ontologies play a critical role in defining, establishing, and sharing the meaning of concepts and relationships between concepts in the form of shared vocabularies. Of special interest is the GeoSpatial Semantic Web

(Egenhofer M. , 2002), which captures the semantics of individual entities, spatial locations and most importantly, spatial relations between these entities. Previous works ((Cohn, Bennett, Gooday, & Gotts, 1997), (Mark & Egenhofer, 1994)) have clearly identified the preferred semantics of spatial relations between geometries through human interaction experiments. This research has laid the ground work for natural language expressions between spatial entities but as formulated between generic spatial primitives (regions, lines, points). For example in the Figure 1.1 , A and B are abstract representations for two spatial entities. Both the (Region Connection Calculus (RCC) (Randell, Cui, & Cohn, 1992) and 9-intersection methods (Egenhofer & Herring, 1991) provide a formal basis for computing this relationship and attaching a label to it. However once these abstract spatial entities are associated with a specific semantic context, other semantically meaningful relationships may apply. For example, suppose that A and B represent administrative units such as cities and towns and a user has a query about which units are suburbs of another (e.g. Is Sugar Land a suburb of Houston?)

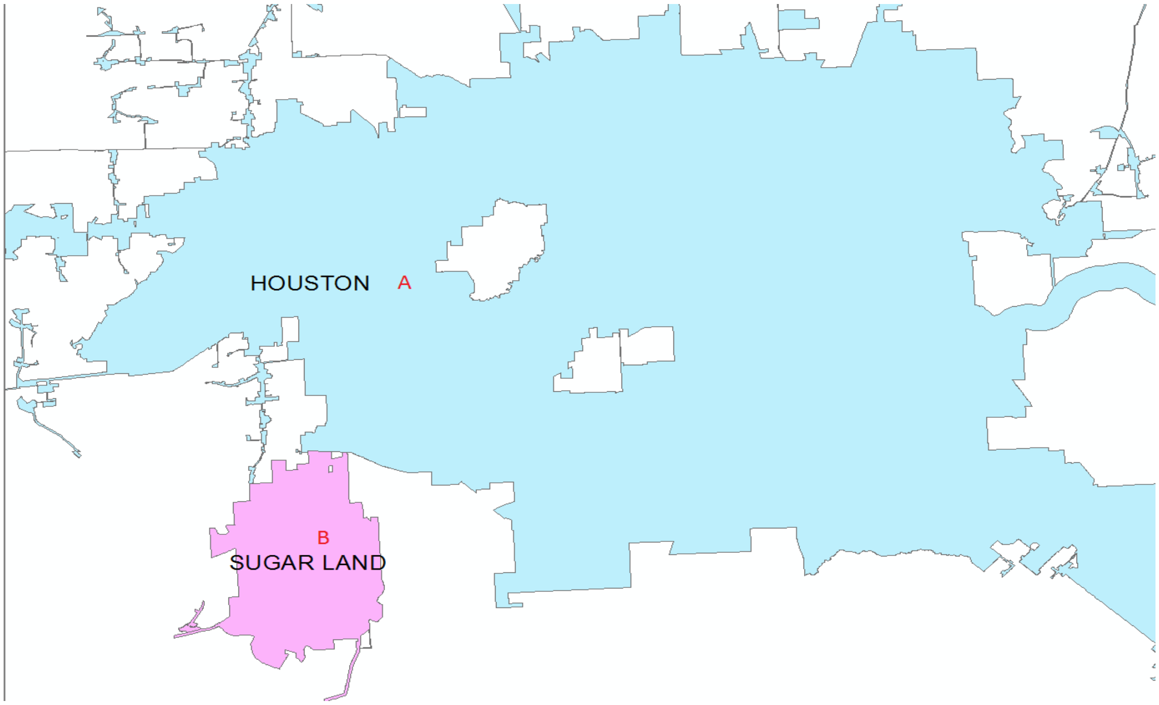


Figure 1.1 Suburbs of Houston

The suburbOf relation implies a spatial relationship but it does not necessarily map directly onto one of the 9-intersection (Egenhofer & Herring, 1991) or RCC relation. We can expect that place names carry semantics. For example many people would recognize El Paso as being a city in Texas and, as a city, we could expect it to have certain relations to other cities and towns (e.g., bordering Ciudad Juarez). The question that this thesis seeks to address is that given the semantics of place names conveyed by the feature type (e.g., city, river, beach), can we identify more domain specific geospatial relationships among feature types that can be used to enhance place name based queries.

Gazetteers currently support the mapping of place names to spatial objects (points, lines, polygons) but few have formally specified relationships between place names. If two place names are mapped to associated spatial objects (footprints), we can obtain the generic spatial relations between their footprints but this relation may miss a semantically richer set of geospatial relations that may exist between place names and in natural language terms on which users may wish to query.

This thesis presents an approach, where place name searches are enhanced by a geo-ontology and a semantically enabled gazetteer. The approach investigates the extension of general spatial relations to domain specific semantically rich concepts and spatial relations. Hydrography is selected as the domain, and the thesis investigates the specification of relations between hydrographic features as functions of spatial relations between their footprints.

1.1 Motivation

Hydrographic feature names almost always require a feature type as part of place name searches such as, “Brazos River,” “Sandy Creek,” or “Addicks Reservoir.” Wikipedia defines a stream as follows:

“A stream is a body of water with a current, confined within a bed and stream banks. Depending on its locale or certain characteristics, a stream may be referred to as a branch, brook, beck, burn, creek, ‘crick’, gill (occasionally ghyll), kill, lick, rill, river, syke, bayou, rivulet, streamage, wash, run or runnel.”

Consider this scenario, the user wants to search for a hydrological feature named Sandy, but the user is not sure whether it is a creek, brook, river, or bayou. There is often an expectation that users are domain experts, which is not always the case. Hence it is essential, to aid the place name searches with a knowledge base, which can specify broader, narrower, and equivalent lexical terms. Another limitation is the format of the geographic name itself. Consider a scenario, where two streams are named "Sandy Creek," one in Texas and another in Wisconsin. In common terms, we identify Sandy Creek, TX or Sandy Creek, WI in order to establish the geographic context we are referring to (Hill & Zheng, 1998). Further, "Sandy Creek" may be represented as a polyline or a polygon, which also adds to the complexity of the search. A spatially referenced gazetteer with geographic footprints and a feature type thesaurus helps to address the two scenarios mentioned above.

A limitation in current gazetteers is that they do not represent underlying feature-feature or feature part_of relationships between the features they represent. Hence place name searches cannot directly accommodate queries for semantically related place names. For example, suppose a user is conducting a bacteria assessment and finds high values at a station in the Brazos River. To explore the problem further the user now wants to find all upstream connected waterbodies. A GIS may realize such a query but a place name query to the Web or to a gazetteer using the name Brazos River cannot directly retrieve features with hydrologic connections to the Brazos River. Although digital gazetteers include spatial footprints (such as points or bounding boxes), the set of features returned based on spatial relationships between these footprints can be incomplete or imprecise based on the level of detail of the footprint. Thus limited spatial semantics narrows the effectiveness of spatial queries and often excludes useful results (Fu, Jones, & Abdelmoty, 2005). However, if the same query can be expanded based on semantic properties, such as *hasTributary* and *hasWaterBody*, that capture hydrologic relations between named features, then components hydrologically connected to the Brazos River can be returned. In order words, search engines need to be augmented with semantically enhanced geospatial searches. Various human experiments have been conducted (Mark & Egenhofer, 1994), (Shariff, Egenhofer, & Mark, 1998) in order to identify semantics of spatial relations between common features and formal representations of these relations have been developed. While these formal spatial relations have associated natural language terms, domain specific semantics may not map to these relations directly or they may map to a combination of these spatial relations. It is very important to note that natural language terms attached to these underlying spatial relations may change depending on the data we are investigating.

The goal in this thesis is to construct a domain-specific ontology that encompasses concepts and their spatial semantics. A computational model for each of the semantic properties defined in the ontology was then implemented. In order to demonstrate the effectiveness of Semantically Enhanced Place Name searches using ontologies, a prototype of a search engine was implemented for searching the developed hydrographic gazetteer.

1.2 Problem Statement

The focus of this research is to develop an ontology-based knowledge discovery and retrieval method. We constructed a geo-spatial domain ontology in the field of surface hydrology, which specifies canonical forms of hydrologic features and captures the semantic meanings of their spatial relationships with one another, if they exist. We also model these spatial relations, in machine-understandable spatial processing methods. Previous works have successfully demonstrated that a geo-spatial hydro ontology can be used to explicitly encode topological operations (Goodwin, Dolbear, & Hart, 2008). In

Goodwin's approach, the topological operators were taken from region connection calculus (RCC8) (Cohn, Bennett, Gooday, & Gotts, 1997). However, this approach captures only the topology of features. The semantic meaning of these spatial relations in the context of hydrology is still implied or undetected. Geospatial searches can be qualitative or quantitative in nature. Egenhofer and Franzosa (Egenhofer & Franzosa, 1991) argue that spatial relationships can be categorized into three groups: (1) topological relations which are invariant under topological transformations of the referenced objects (Egenhofer & Herring, 1990), (2) metric relations in terms of distances and directions (Pequet & Ci-Xiang, 1992), and (3) relations concerning partial or total order of spatial objects as described by prepositions such as "in front of", "behind", "above" and "below" (Herring, 1991). RCC8 and the Egenhofer 9-intersection based spatial operators have formalized representations in Open Geospatial Consortium (OGC) standards (GEOSPARQL) (Perry & Herring, 2011) and are widely adopted in commercial GIS products. In order to take advantage of these operators in place name searches, context based semantic mapping to a combination of these operators is investigated in this thesis.

The hypothesis of this thesis is that *Semantic Feature-Feature relationships are derivable from spatial geometry relations subject to domain constraints.* For example, if Sugar Land is to be considered as a 'SuburbOf' Houston, it has to satisfy the geometry relation of being adjacent to another polygon representing Houston Metro Area. However this spatial condition might apply to other features such as a large park which may lie adjacent to Houston, but do not satisfy other domain constraints, such as being a town for a suburb. In a hydrological context, a stream is considered a tributary if it joins the main stem of a river or if it is a tributary of another stream that flows into the river. By explicitly representing these hydro-relations, queries such as, What are the tributaries of Brazos River? ,and What are the rivers that flow into the Gulf of Mexico? can be supported directly instead of relying on generic spatial relations. The National Hydrography Dataset (NHD) for the regions of Maine are used for the development, implementation, and testing of the proposed geospatial hydro ontology and gazetteer.

The following steps outline, the methodology adopted in this thesis.

- Identify a set of prototypical hydrologic features (stream, lake spring, wetland, ocean).
- Identity semantic relationships between the hydrologic feature types (stream-stream, stream-lake, lake-lake, etc.).
- Develop an ontology of the prototypical feature types and their relationships.
- Investigate mappings between specified semantic feature–feature relationships and possible supporting spatial relations.
- Implement spatial operations to derive the semantic feature-feature relationships from footprint spatial relationships.
- Test the implemented gazetteer and semantic relationships with a set of competency questions for retrieving features.

The outcome of this methodology is a semantically enhanced gazetteer of named hydrologic features, along with their hydrological relationships indicated by semantics. This repository can now be used to query based on semantic concepts as well as place names.

1.3 Research Questions

As the outcome of this research, we intend to answer the following questions.

- Can spatial semantics be used to improve the completeness of place name search results? For example, a search for “Brazos River Basin” returns all tributaries, isolated networks and water bodies in the Brazos River Basin area.
- How do feature-feature semantic relationships map to topological and mereological relationships For example, a search for “Tributaries of Brazos River” returns all tributaries that are connected to the main stem of the Brazos River and are also a hydrological part of the river system.
- Do spatial semantics provide better query disambiguation and enable complex spatial analysis? For example a search for “What are the sources of a River X”, return all the sources of the River X, which can be headwater, springs, or lakes irrespective of the feature type.

1.4 Scope of Study

The scope of this research is to identify, describe and implement a geo-spatial hydro ontology and gazetteer that cover hydrological concepts and semantic relationships. This research also builds a prototype search application which supports place name searches and searches on relationships between named places. This research focuses on semantic refinement of the topological connectivity relations prevalent in a hydrologic network comprised of streams, lakes, springs, and some hydrographic structures such as dams and bridges.

1.5 Organization of the Thesis

The remainder of the thesis is organized as follows. Chapter 2 covers a review of previous works and supporting literature. Chapter 3 discusses the GazOntology developed to model the digital representations of geographic features based on ISO standards. Chapter 4 identifies prototypical hydrological features and elaborates on hydrological relationships between these features along with the formal representation of hydrological features and their relationship in a hydro ontology. Chapter 5 examines footprints for these feature types and spatial relationships between them as modelled in the NHD Datasets and describes the spatial analysis methods used to generate instances for the HydroGazetteer. Chapter 6 describes implementation of a hydrological gazetteer based on the ontology, semantic web technologies and the prototype web application that interacts with the HydroGazetteer. Chapter 7 summarizes conclusions and describes future work.

CHAPTER 2

LITERATURE REVIEW

This thesis proposes an ontological approach for a semantically enhanced gazetteer for place name searches. This chapter presents background literature relevant to this thesis topic. This chapter covers place name searches and various approaches proposed in previous works to improve place name queries. It reviews existing works on the use of geo-ontologies for efficient information retrieval. Spatial relations are important components of geospatial searches and thus relevant topics on spatial relations are presented. As hydrology is the chosen domain for implementation and testing, related research on domain ontologies in the field of hydrology are briefly reviewed. Semantic web technologies including RDF, RDFS, SPARQL and GEOSPARL as the implementation platforms are also reviewed.

2.1 Place Name Searches

People query for geographic information in the form of place names, addresses, and zip codes, often with the inclusion of spatial qualifiers such as direction (e.g., North of Houston) and proximity (e.g., near Houston). Often these queries retrieve geometric coordinates among other spatial and non-spatial attributes of the subject of interest. Such geographic information retrieval typically involves translating the user query into geographic coordinates through some form of geo-referencing. Informal geo-referencing covers situations where geographic locations are implied by the use of place names, administrative hierarchies, and place types. This section reviews previous work related to extending place name searches with gazetteers and associated ontologies.

Digital gazetteers contain structured information about named places that have a particular geographic location (Goodchild & Hill, 2008). The generally accepted requirement for a digital gazetteer is to hold place descriptions represented as a tuple (Name, Footprint, Type or category). Thesauri (e.g., Getty thesaurus of Geographic Names), Gazetteers (ADL Gazetteer), and metadata structures (e.g. MARC) are informal geo-referencing sources where geographic locations are stored as one of the many components used to identify a specific entity. The Alexandria Digital Library Gazetteer (Hill & Zheng, 1998) is one of the early and well-recognized digital gazetteers created through the combination of the Geographic Names Information System (GNIS) and the Geographic Names Processing System databases, both from US federal-government agencies. Another recently developed on-line gazetteer, GeoNames, contains over 10 million geographic features covering the world and categorized into sets of feature classes and subclasses (http://www.geonames.org/about.html). Some of the data sources used by the GeoNames gazetteer include The National Geospatial-Intelligence Agency's and the U.S. Board on Geographic Names (most names except US and CA), and U.S. Geological Survey Geographic Names Information System (names in US). GeoNames also complements its database with geotagged information from Wikipedia.

The role of digital gazetteers in enhancing place name search has been well researched (Goodchild & Hill, 2008). Work by Jones et al. (Jones, Alani, & Tudhope, 2001) started with an objective to implement procedures that match a given place name to named places that are equivalent or similar in geographic location. They developed a prototype Ontology, OASIS (Ontologically Augmented Spatial Information System), with a mix of qualitative and quantitative spatial data including topological relations and approximated point coordinate data representing the centroid of a feature. A *Place* concept was implemented as a type of *Geographic Concept* and a place was modelled to include multiple places through the topological relationships meet and part of. OASIS also contained cultural information about historical places that were classified using terms from the Art and Architecture Thesaurus and linked to a thesaurus of geographic names.

Gazetteer concepts for information retrieval in the web were enhanced by (Jones, Abdelmoty, & Fu., 2003) using a base schema for a geographical ontology that supported multiple footprints for each feature. Spatial relations were supported, but were limited to beside, near, overlap, inside, disjoint, and touch. Each of the spatial relations was extended with synonymous spatial relations terms, for example, the spatial relation *beside* includes two synonymous relations *alongside* and *next-to*. Similarly spatial relation *touch* includes a number of synonymous relations, such as *adjacent, on the boundary of, next, side by side, close.*

As a part of the SPIRIT project, Fu et al. (2005) presented a geo-ontology of places and employed four similarity measures to identify geographical places by place name, place type, footprint, and a geographical hierarchy to assist spatial search in the web (Fu, Jones, & Abdelmoty, 2005).

Janowicz and Keßler (Janowicz & Keßler, 2008) investigated the role of a feature type ontology in improving gazetteer interaction. Their work demonstrated that the development of shared feature type ontology can support similarity assessment through subsumption relationships that no longer require users to know what is meant by a specific feature type.

Wu and Winter (2009) presented an approach to measure the similarity between gazetteer instances and place names at three levels: string similarity where the place name is matched with gazetteer instances to see if there is an exact match. If this step did not produce the desired result, ontological similarity was considered where the feature type is matched using an ontology of feature types. The resulting set is then run through a spatial similarity process with an assumption that the location of the user is known and feature being searched for is in close proximity to the user (Wu & Winter, 2009).

A number of researchers have addressed the problem of ambiguity in place names. Hasting (2008) addressed the resolution of ambiguous place names in the conflation of multiple gazetteer data (Hastings, 2008). Overell (2011) discussed the issues of place name ambiguities and highlighted the exploitation of topological and geographic relations between locations in solving the issue of place name ambiguities (Overell, 2011). A few gazetteer implementations have incorporated explicit relations between named features to varying degrees. The gazetteer model for the Alexandria Digital Library specified a generic isRelatedTo relation between features with the intent that this could be specialized over time (Hill, 2000). The Getty Thesaurus includes a spatial containment hierarchy among named places (Getty Research Institute 2014). The ontology underlying the GeoNames gazetteer includes three relationships for connecting features: a "children" relation, that links administrative sub-divisions to countries, and "neighbor," and "nearBy" relations that connect features in close proximity. For the most part, however, gazetteers have remained largely flat structures with named places and features as isolated unconnected instances. Spatial relations between the feature footprints can be obtained on the fly, but these may not always translate to the semantic relations between features desired and sought by users. This thesis seeks to address this gap by investigating an approach to incorporate semantic relations among named features. The identification of semantic relations between features uses the underlying topological relations between feature footprints as a starting point in combination with characteristics of the real world features. The next section provides a background on spatial relations between geometric primitives that provides one basis for deriving feature-feature semantic relations.

2.2 Spatial Relations

Spatial relations, in combination with place names provide an important addition for supporting more expressive geospatial searches. Three classes of spatial relations have been identified based on different spatial concepts (Pullar & Egenhofer, 1988) (Shariff, Egenhofer, & Mark, 1998).

(1) Topological relations are invariant to topological transformations such as translation, rotation and scaling (Egenhofer M. , 1989). Example terms include *neighbor, overlap,* and *disjoint*.

(2) Spatial order relations and strict order relationships rely upon the definition of order. They are invariant under translation and scaling but subject to change under rotation. Each order relation has a converse relation. For example an object A is *behind* object B based on the order definition *preference.* The converse of this relation is object B is *in front of* object A.

(3) Distance relations express measurements that reflect the concept of a metric and change under scaling but are invariant under translation and rotation. Example terms

include *'near', 'within 5 mile radius of I-95'*. Distance relations may be expressed qualitatively or quantitatively and are often used to refine disjoint relationships (Mark, 1999). Three different formal approaches for defining spatial relations exist in the literature. The first method by Peuquet makes use of distance and direction primitives in combination with logical connectors AND, OR and NOT (Peuquet, 1986). The relation disjoint(A,B) is represented by the constraint that the distance from any point on A to any point on B is greater than 0. However this approach does not consider inclusion or containment, unless negative values for distance are introduced (Egenhofer M. , 1989) .

The point-set approach (Egenhofer & Franzosa, 1991) describes binary topological relationships by comparing points of two objects with conventional set operators. For example, the relation inside(x,y) is expressed as points(x) ⊆ points(y). While equality, inclusion and intersection can be described in this approach, neighborhood relations cannot be described.

The third approach represents relationships based upon the intersection of the boundary, interior, and exterior of two objects to be compared and distinguishes them based on their intersections only (Egenhofer M. , 1989). The advantage of this method is that it describes topological relations purely based on topological properties. Topological information is qualitative and does not consider direction or distance measures. For example, two objects are said to be neighbors if they share a common boundary and the length or area of the common boundary is immaterial in order to determine the neighborhood property. While spatial terms can have different meanings depending on the application domain, all spatial relations are based upon the fundamental geometric properties of the objects which are represented as points, lines or regions (Egenhofer & Herring, 1991).

2.2.1 RCC8 Relations

Randell et al. (Randell, Cui, & Cohn, 1992) describe interval logic for reasoning about space using a simple ontology that defines functions and relations for expressing and reasoning over spatial regions. This logic is referred to as Region Connection Calculus (RCC). The ontological primitives of this theory include physical objects, regions, and other sets of entities. The basic part of the formalism assumes one primitive relation, C(x,y) read as 'x connects with y.' This connection relation is both symmetric and reflexive. Based on the connection relationship, other relations can be defined, such as parthood, P. For example, P(x,y), means that x is a part of y as long as anything connected to x is also connected to y. A subset of RCC, RCC8, defines eight mutually exhaustive pairwise disjoint relations, which can be used to define the rest of the relations in RCC. These eight base relations are:

1. DC(x, y) (x is disconnected from y)

2. x = y (x is identical with y)

3. PO(x,y) (x partially overlaps y)

4. EC(x,y) (x is externally connected with y)

5. TPP(x,y) (x is a tangential proper part of y)

6. NTPP(x,y) (x is a non-tangential proper part of y)

7. TPPi(x,y) (y is a tangential proper part of x)

8. NTPPi(x,y) (y is a non-tangential proper part of x)

2.2.2 Egenhofer Relations

Egenhofer's intersection models (Egenhofer, Sharma, & Mark, 1993) have been investigated for points, lines, and regions and are extended based on the different dimensions involved, in contrast to the RCC theory, where the authors do not explicitly define what comprises a region and all geometries are considered as spatial primitives. This thesis considers primarily the topological relations between hydrological features, which may be represented with point, line, and region geometries. Hence, this section reviews two primary models used for binary topological relations, the 4-intersection and 9-intersection, which is an extension of the 4- intersection model.

The intersection models assume a spatial object A is identified with parts interior (A°), boundary (∂A) and exterior (A^-). The 4-intersection model, is represented as a 2x2 matrix between two spatial objects A and B based on the intersections of A's boundary (∂A) and interior (A°) with B's boundary (∂B) and interior (B°). The 9-intersection model is an extension that considers the location of each interior and boundary with respect to the other object's exterior. It is represented as a 3x3 matrix based on the intersections of A's interior (A°), boundary (∂A), and exterior (A^-). with B's interior (B°), boundary (∂B) and exterior (B^-). This intersection model has been investigated for region-region, region-line (Shariff, Egenhofer, & Mark, 1998) line-line relations in R^1 (Pullar & Egenhofer, 1988), and line-line in R^2 (Clementini, Felice, & Oosterom., 1993) (Mark & Egenhofer, 1994) and directed line (Kurata & Egenhofer, 2006) relations.

2.2.2.1 9-Intersection between Directed Line Segments

Since hydrologic feature relations are characterized by flow direction, important relationships are those between directed line segments. A directed line segment is defined to consist of three parts: two distinct points and a non-self-intersecting line segment that connects the two points, and a direction which establishes the starting and the end points (Kurata & Egenhofer, 2006). Since, directed line segments are geometrically similar to arrows, topological relations between two directed line segments are captured like the topological relations between arrow symbols (Kurata & Egenhofer, 2006). The three parts of the directed line segment are identified as the head slot, tail slot, and body slot of the arrow diagram as shown in Figure 2.1.

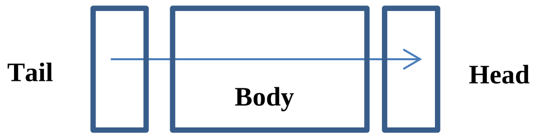


Figure 2.1 Head- Body-Tail of an Arrow Representation

Considering two directed line segments X and Y, their topological relationships are based on the intersections between the interiors and two ordered boundaries (head and tail). In terms of the 9-intersection model and using components of the head-body-tail of the directed line segments, the set of possible relations between directed line segments are summarized in Equation 2.1 (Kurata & Egenhofer, 2006).

$$M_I(A,B) = \begin{bmatrix} \partial_{tail}A \cap \partial_{tail}B & \partial_{tail}A \cap B^\circ & \partial_{tail}A \cap \partial_{head}B \\ A^\circ \cap \partial_{tail}B & A^\circ \cap B^\circ & A^\circ \cap \partial_{tail}B \\ \partial_{head}A \cap \partial_{tail}B & \partial_{head}A \cap B^\circ & \partial_{head}A \cap \partial_{head}B \end{bmatrix}$$

Equation 2.1 Head Body Tail Intersection of Directed Line Segments

Of the possible configurations of this 3x3 matrix with empty and non-empty relations, 68 matrices were identified to have valid geometric interpretations. These are referred to as topological relations classes (TR- Classes) and given names based on the name primitives; split, diverge, precede, divergedBy, cross, mergedBy, follow, merge, and meet assigned to the relationship types (Equation 2.2 (Kurata & Egenhofer, 2006)). Because digital river or stream representations are typically represented by directed line segments, these relations among directed line segments provide the basis for deriving the semantic relationships between streams.

$$\begin{pmatrix} split & diverge & preceed \\ divergedBy & cross/touch & mergedBy \\ follow & merge & meet \end{pmatrix}$$

Equation 2.2 Relations between Directed Line Segments

2.2.2.2 9-Intersection between Directed Line Segment and Region.

Movement of an agent with respect to an area, such as entering, leaving, and passing through, is modelled as a spatial relation between a directed line segment and a region. Topological relations between a directed line segment and a region capture the possible patterns of an agent's movement with respect to a region. By considering a directed line segment D and region R, possible topological relations are captured through geometric intersections of R's three respective topological parts (interior, boundary, and exterior) with respect to D's boundary distinguished into intersections with respect to D's starting and ending points (Kurata & Egenhofer, 2007). The following basic qualitative conditions were considered to represent all movement patterns: (1) starting from interior, (2) starting from boundary,(3) starting from exterior, (4) crossing boundary, (5) ending at interior, (6) ending at boundary, (7) ending at exterior and (8) crossing/ touching boundary (Kurata & Egenhofer, 2007). In this thesis these relations are used to identify flow relations between streams represented as directed line segments and lakes or other water bodies represented as regions.

2.3 Semantic Web Technologies and Linked Data

The vast majority of the resources available on the web are HTML pages which present machine-readable content in human readable format. However, a search algorithm cannot understand the structure or semantics of a resource, unless web pages expose enough information about the type of the resource or information about other resources that are interconnected with the resource. The idea of the semantic web was first introduced by Tim Berners-Lee (Berners-Lee, Hendler, & Lassila, 2001) at the first World Wide Web consortium in 1994 (Berners-Lee & Cailliau, 1994) (Shadbolt, Hall, & Berners-Lee, 2006). The Semantic Web is a Web of actionable information, derived from data through a semantic theory for interpreting the symbols. The semantic theory provides an account of meaning in which the logical connections among terms establish interoperability of systems. Berners-Lee et al. describe linked data as simply using the Web to create typed links between data from different sources (Bizer, Heath, & Berners-Lee, 2009). To achieve this, a framework for modelling and structuring data and information is needed. The Resource Description Framework (RDF) is the model for describing and connecting information in the sematic web and SPARQL (Prud'Hommeaux & Seaborne, 2008) is the query language that is used to search and retrieve information from linked data in the semantic web.

2.3.1 RDF

The Resource Description Framework (RDF) (Manola, Miller, & McBride, 2001) is an XML based approach to represent semantic information in the web recommended by the World Wide Web Consortium (W3C). RDF is designed to store information about a resource in a machine readable way and also preserves its meaning by storing the type of information in a standardized format along with the information itself. Every resource represented in RDF may be a reference to an actual entity in the real world. Universal Resource Identifiers (URI) and now Internationalized Resource Identifiers (IRIs) are used to uniquely identify an individual resource. The RDF primer document (Manola, Miller, & McBride, 2001) introduces an abstract format for representing RDF statements. Each RDF statement contains a *subject*, *predicate*, *and object.*

An RDF statement relates the subject with the object by means of the predicate. Subject and objects are resources and the predicate is the relationship between the two resources and may also be a resource itself. The relationship is phrased in a directional way (from subject to object) and is referred to in RDF as a *property*. Because RDF statements consist of three elements they are called *triples*. A resource can have attributes which in turn can have values. RDF can also use other data types such as integers, dates, and string literals as values of the properties.

For example, the following RDF statements or triples present some information about Bob.

<Bob> <is a> <person>.

<Bob> <is a friend of> <Alice>.

<Bob> <birthdate> <the 4th of July 1999>.

<Bob> <has Address> <228 Glenridge Forest>

Some of the serialization formats for RDF recommended by the W3C are Turtle and TriG, JSON-LD (JSON based), RDFa (for HTML embedding), N-Triples and N-Quads (line-based exchange formats), RDF/XML (http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/).

Every resource and property is assigned an International Resource Identifier (IRI), a generalization of a URI (Uniform Resource Identifier), allowing non-ASCII characters to be used in the IRI character string. By associating a resource with an IRI in the semantic web, anybody can link and refer to it. As shown in the example below, RDF assigns a specific IRI to resources which may be classes, objects or relations between objects.

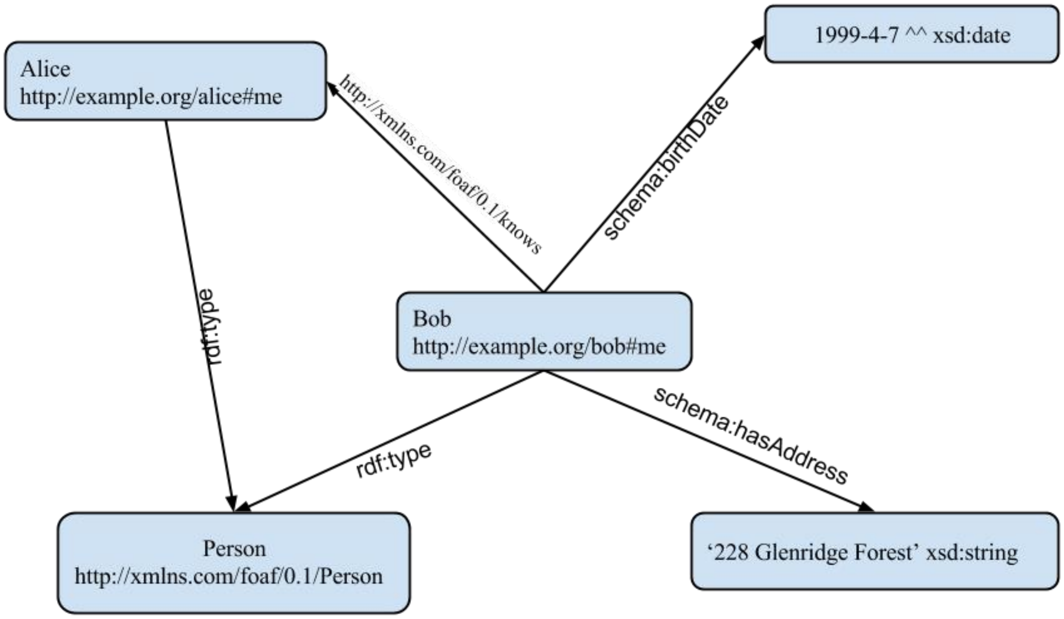

Alice
http://example.org/alice#me
1999-4-7 ^^ xsd:date
http://xmlns.com/foaf/0.1/knows
schema:birthDate
rdf:type
Bob
http://example.org/bob#me
schema:hasAddress
rdf:type
Person
http://xmlns.com/foaf/0.1/Person
'228 Glenridge Forest' xsd:string

Figure 2.2 RDF Graph Representation

2.3.2 RDFS

The RDF data model provides a way to make statements about resources. This data model does not make any assumptions about what resource IRIs stand for. In practice, RDF is typically used in combination with vocabularies or other conventions that provide semantic information about these resources. RDF Schema language (RDFS) provides a richer set of vocabularies to allow the definition of semantic characteristics of RDF data. For example, one can state that the IRI http://www.example.org/knows can be used as a property and that the subjects and objects of http://www.example.org/knows triples must be resources of class http://www.example.org/Person.

The main modelling constructs provided by RDF Schema are summarized in Table 2.1 which is adapted from the RDF primer document by W3C.

Construct	Syntactic form	Description
Class (a class)	**C** rdf:type rdfs:Class	**C** (a resource) is an RDF class
Property (a class)	**P** rdf:type rdf:Property	**P** (a resource) is an RDF property
type (a property)	**I** rdf:type **C**	**I** (a resource) is an instance of **C** (a class)
subclassOf (a property)	**C1** rdfs:subClassOf **C2**	**C1** (a class) is a subclass of **C2** (a class)
subPropertyOf (a property)	**P1** rdfs:subPropertyOf **P2**	**P1** (a property) is a sub-property of **P2** (a property)
domain (a property)	**P** rdfs:domain **C**	domain of **P** (a property) is **C** (a class)
range (a property)	**P** rdfs:range **C**	range of **P** (a property) is **C** (a class)

Table 2.1 RDFS Properties

RDF Schema uses the notion of **class** to specify categories that can be used to classify resources. The relation between an instance and its class is stated through the **type** property. With RDF Schema one can create hierarchies of classes and sub-classes and of properties and sub-properties. Type restrictions on the subjects and objects of particular triples can be defined through **domain** and **range** restrictions. An example of a domain restriction was given above: subjects of ‘knows’ triples should be of class ‘Person’.

2.3.3 OWL – Web Ontology Language

The Semantic Web expresses information with explicit meaning, so that machines can automatically process and integrate information available on the web. RDF provides the schema to represent resources and relationships between them. RDFS adds rich vocabularies to represent classes and properties. However, Ontology is required to formally describe the meaning of the terminology used in web documents. If the expectation of the processing algorithm is to support reasoning from the web documents, the language in the documents has to express semantics in such a way that it accommodates the RDF schema model. OWL is a web ontology language, which is the W3C recommendation for authoring ontologies. OWL extends DAML + OIL ontology language. OWL adds more vocabulary for describing properties, classes and relations between classes such as disjointness, cardinality, equality, and characteristics of properties such as symmetry. OWL provides three increasingly expressive sublanguages: *OWL Lite* that supports uses primarily needing a classification hierarchy and simple constraints. *OWL Description Logics* supports users with need for maximum expressiveness with computational completeness. *OWL Full* supports users with the need for maximum expressiveness and represented as RDF with no computational guarantees.

OWL makes use of RDFS features discussed in the section above and XML language constructs to represent data types along with the following property characteristics: (1) Object Property, (2) Data type Property, (3) InverseOf, (4) Transitive Property, (5) Symmetric property, (6) Functional Property, and (7) Inverse Functional Property.

2.3.4 SPARQL

Once a repository of RDF statements, called a triple store, is created, a query language is needed to get meaningful results and inferences. SPARQL is analogous to SQL in RDBMS and is used to query and retrieve results from a triple store. SPARQL (Prud'Hommeaux & Seaborne, 2008) is the official W3C recommendation for querying RDF data. The syntax and concepts presented here are based on the recommendations made by the SPARQL Working Group. SPARQL standards became available in January

2008 (http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/) and SPARQL 1.1 is the most recent document from W3C (http://www.w3.org/TR/sparql11-query/). SPARQL querying is based on graph pattern matching. Triple patterns contain variables in any part – subject, predicate or object, and results are obtained by matching the non-variable part of the triple pattern against triples in a triple store.

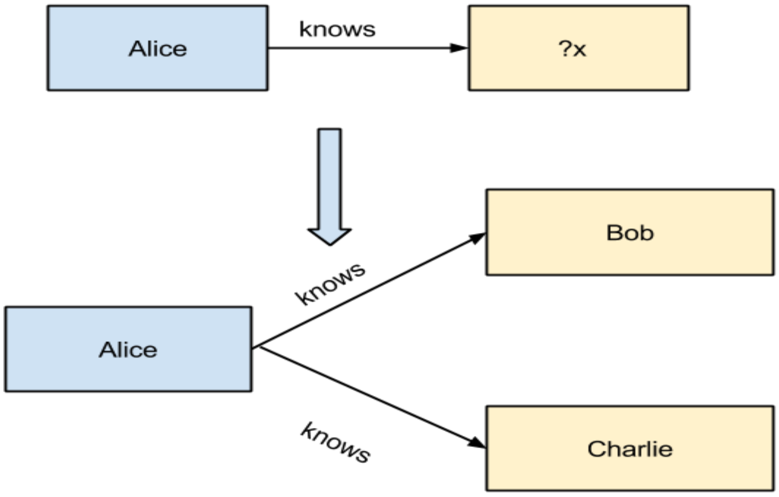


Figure 2.3 Graph Pattern Matching

Basic graph patterns are sets of triple patterns. To create complex patterns, graph patterns can be connected and manipulated with a full stop after each triple pattern. Semantically, this is regarded as the conjunction of each included tuple. Figure 2.3 shows an example of graph pattern matching.

Alice knows ?x . will return Alice knows Bob and Alice knows Charlie since both the subject and the predicate matches in this set of tuples. Group graph patterns can be constructed by combining variables from different graphs such as

{Alice knows ?x . ?x likes ‘Mona Lisa’. }

This query will return all the persons that Alice knows who like ‘Mona Lisa’. The results can be further restricted by the usage of a FILTER key word in combination with the respective variable. SPARQL provides keywords similar to SQL including SELECT and WHERE clauses, to select a set of triples that satisfy the criteria specified in the where clause. The CONSTRUCT query type is used to create a new RDF graph from the data given by the pattern in the where clause. The ASK query matches the pattern given after ASK with the graph and returns a Boolean yes or no, if there is at least one match or zero, respectively.

2.3.5 GEOSPARQL

GEOSPARQL is the spatial extension of the SPARQL query language and defines a core set of classes for representing geospatial information in the web as RDF statements and for performing spatial computations. GEOSPARQL is comprised of a core component that defines the top level RDFS/OWL classes, a topology vocabulary that helps in identifying the spatial relationships between two spatial objects, a geometry component that defines RDFS data types for serializing geometry data (Terse RDF triple language, RDF/XML), a geometry topology component that defines topological query functions for geometry objects (Egenhofer and RCC8 relations are implemented), a query rewrite extension and RDFS entailment section (Perry & Herring, 2012).

Other schemes for encoding simple geometry data in RDF have been proposed. The W3C Basic Geo vocabulary (http://www.w3.org/2003/01/geo/) is one popular vocabulary. These simple vocabularies have limitations such as only point geometries are supported and the inability to specify different datum and coordinate systems, and are therefore not used in GEOSPARQL. Most existing geometry data encoded using these vocabularies can easily be converted into GEOSPARQL representations as Well Know Text representation and GML representation of geometries are supported in GEOSPARQL.

2.4 Domain Ontology

The idea of formal ontology was first suggested by Edmund Husserl, who drew the distinction between *formal logic* and *formal ontology* (Smith, 1998). Formal logic deals with the interconnections of truth with inference relations and their validity. Formal Ontology deals with the interconnections of physical entities with objects and properties. Husserl's formal ontology is based on three categories: (1) theory of mereology (part-whole), (2) theory of dependence and (3) theory of topology (boundary, continuity and contact). Formal Ontology refers to an ontology as a particular system of representing reality in a philosophical sense. (Guarino, Semantic matching: Formal ontological distinctions for information organization, extraction, and integration., 1997) classified formal ontologies based on two dimensions a) level of detail and b) level of dependence on a particular context. Based on the level of detail, ontologies can be classified as coarser and finer ontologies. Finer ontologies include more specialized vocabularies when compared to the coarser ontologies. Guarino's classification refers to an ontology as an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words (Guarino, 1998).The domain ontology development in this thesis is positioned with a 3-layered architecture of ontologies that distinguish: top-level ontologies, domain and task ontologies, and application ontologies.. Top-level ontologies describe general concepts such as space, time, object, process, and events, which are common to all domains and remain domain neutral and thereby facilitate semantic interoperability between other ontologies. Domain ontologies and task ontologies describe concepts, relations and vocabulary specific to a given domain. These concepts are expressed as specializations of entities defined in top-level ontologies. Application ontologies combine and extend the domain ontology concepts and correspond to the different roles played by domain entities (Guarino, Semantic matching: Formal ontological distinctions for information organization, extraction, and integration., 1997).

Various upper-level ontologies have been implemented, such as Basic Formal Ontology (Smith, 1998), General Formal Ontology, DOLCE etc. In this thesis, DOLCE classes are extended to represent real world hydrologic features.

2.4.1 Domain Ontology Developments related to Hydrology

(Schwering, 2004) explores the use of spatial relations for the ad-hoc integration of geo-information. This paper proposes spatial relations as an additional way to calculate similarities and demonstrates that other relations apart from hyponyms and meronyms are necessary to calculate similarity. This work modelled river, dam, carse, watermeadow, and flood basin as specializations of a Surface water body class. The relation ‘next-to’ exists between river and other concepts. According to the ecological data sources, floodplains are periodically flooded areas next to rivers. Potentially related concepts can be investigated by searching the spatial neighborhood (next-to relations) of concept River. Schwering and Raubal extended this work, by creating a shared vocabulary of spatial relations such as along, connected to, in, end in, end at, end just inside, end near and near / very near (Schwering & Raubal, 2005).

The Ordnance Survey maintains a continuously updated database of the topography of Great Britain, including around 440 million geographic features, such as forests, roads and rivers down to individual houses, and garden plots. In order to take advantage of the semantics of this large geospatial dataset, a hydrology ontology containing 301 classes and 162 properties was constructed as a subset of the topography ontology (Dolbear, Hart, & Goodwin, 2006).

Further, (Goodwin, Dolbear, & Hart, 2008) investigated the use of linked data in the web to represent the topographic data of Great Britain. The authors created an administrative ontology as a part of this investigation to represent the administrative areas of Great Britain. An administrative geography gazetteer was developed with the purpose of introducing a vernacular gazetteer at a later date, with explicit linking between the two. RDF datasets were created to represent the administrative ontology along with a topological hierarchy for the four topological relations *Completely spatially contains, Tangentially spatially contains, Borders, Spatially equivalent* from the RCC8 calculus.

Below is an example of statements representing Southampton.

<http://os.rkbexplorer.com/id/osr7000000000037256>

rdfs:label “The City of Southampton”;

rdf:type UnitaryAuthority;

admingeo:hasOfficialName “The City of Southampton”;

admingeo:hasVernacularName “Southampton”;

admingeo:hasBoundaryLineName “City of Southampton (B)”;

admingeo:hasCensusCode “00MS”;

admingeo:hasArea “373814.131”;

admingeo:borders *http://os.rkbexplorer.com/id/osr7000000000001776*

The USGS has also embarked on the development of linked data and the semantic web to represent its extensive data sources in a machine readable form and make them available to a wide variety of applications benefiting scientific communities. As a pilot project, the USGS investigated nine test areas which included 6 sub watersheds and 3 urban areas along with the eight standard The National Map (TNM) datasets to research the possibilities of converting its vast geospatial data to RDF.

Weigand (Wiegand, 2010) presented a non-traditional use of OWL ontologies for query expansion using subsumption relationships for a specific context. She presented an approach for modelling features within a domain based on specific criteria such as ' potential land for production of bio-fuel'. Her case study used the National Land Cover Dataset from TNM. Instead of modelling the full land cover domain, Weigand proposed to create a specialized ontology only for the available land that has potential for bio-fuel production. She also presented examples of querying for a feature across different layers within TNM. By taking a feature-based approach and using domain ontologies, she successfully demonstrated querying features across different domains. This study investigated queries, such as “Find all vehicle transit objects over water bodies that are part of the Wisconsin River System where vehicle_Transit belongs to Transportation Ontology, and Water bodies, Wisconsin River system belongs to HydroOntology." This study makes the case that modelling spatial relations semantically allows users to make queries and draw inferences over spatial data without complex analysis methods.

Usery et al. (Usery & Varanka, 2012) presented a case study where the authors explored methods to build semantics for topographic spatial data using taxonomy and ontology and converting point, vector and raster data to RDF triple stores for semantic access, query and retrieval including geometry. The approach connects semantics with geometry on vector and raster pixels allowing for displaying features in a map. Vocabularies of topographic features to be represented as subjects or objects in the triple store were developed from standard feature list sources of topographic data. Point, polyline and raster data were converted from ArcGIS native geodatabase features to GML OGC standards. This study answered competency questions like "What features intersect any feature with NHD reach code X? , What are the tributaries of River X?". However the ability to query based on semantic prepositions that reflect the underlying topology is limited. One limitation of this work is that although the geometry of features can be retrieved, only the RCC8 and Egenhofer relations between 2d regions are widely implemented and available for spatial analysis. Usery et al. demonstrated that direct tributaries of River X can be retrieved, but indirect tributaries of River X cannot be determined because indirect tributaries are not directly connected to River X and hence will not satisfy the primitive spatial relations.

A surface hydrology ontology design pattern was recently developed at Geo-Vocabulary Camp (GeoVocampDC2013 http://vocamp.org/wiki/ GeoVoCampDC2013) by domain experts as a minimalized domain ontology that can be further aligned with other foundational ontologies to represent surface water features. This design pattern differentiates surface waters from landforms that hold surface water as a Wet module and Dry module respectively. The surface water domain ontology developed in this thesis aligns with the essential features of this design pattern.

(Klien & Lutz, 2005) presented an approach to automate annotation of geospatial features by characterizing the features based on their spatial relationships with other features. The automated annotation scheme identifies a geographical feature “floodplains” if it satisfies a set of criteria that is L is adjacent to a river, L is flat and L is at most 2m higher than the adjacent river. The first condition “adjacentTo” is implemented as a set of GIS operations. This approach makes use of a reference dataset which is a well- known geometry of rivers and a dataset that needs to be annotated which contains a set of floodplains. A classification algorithm applies the spatial relation adjacentTo to the dataset that needs to be annotated along with the other two criteria and features that satisfy all the above criteria are annotated as FloodPlains. This paper demonstrated that spatial relations such as topology, direction and distance between geographic entities can be used to automatically annotate geospatial concepts in geospatial datasets. However, the spatial relation adjacentTo can have different spatial interpretations depending on the context and hence an exhaustive list of spatial operations needs to be implemented for the same spatial preposition 'adjacent'.

(Vilches-Blázquez & Luis M., 2010) demonstrated the use of multiple ontologies and geospatial data from different Spanish public data sources along with demographical data for integration and searching. Several data sources containing information about

Administrative Units, Hydrography and Statistical Units were identified from the INSPIRE project. SVACO, HydrOntology, WGS84 and GML ontologies were used to present geographical features in an interactive web map.

Hydrologic information is generated and published by many government, research, commercial and citizen groups around the world. In order to facilitate interoperability between heterogeneous data with little semantic consensus, OGC adopted WaterML 2.0 as an OGC standard for hydrologic data. General characteristics of WaterML include semantics of hydrological time series to support correct interpretation of time series data and properties (Taylor, 2012).

RiverML 0.2 is a prototype transfer language for storing river terrain geometries and river flow models. RiverML is designed to support interoperability between terrain processing software like ArcGIS, hydrologic calculation software and hydraulic software (Jackson, 2014). This standard is designed to meet the needs of the CUAHSI HydroShare project.

2.5 Summary

This chapter reviewed existing works on digital gazetteers and the use of ontologies in place name searches. It further discussed some of the semantic web technologies including OWL, SPARQL, RDF, RDFS and GEOSPARQL that are used in this work to model and query hydrological features. The chapter also briefly reviewed the 9 intersection model for spatial relations between directed line-segments and between directed line segments and regions, as these are most pertinent to a hydrological network context. Lastly the chapter presented a review of existing literature relevant to the hydrological domain and ontologies developed to represent, query and retrieve hydrological features.

CHAPTER 3

GAZETTEER ONTOLOGY

Representing geographic information in a machine-searchable format is important for efficient query execution and meaningful result generation. Such spatial representations must support data interoperability for querying across different data sources and linking data in related domains. This chapter discusses spatial representations in gazetteers and the gazetteer ontology, which is one of the ontological schemas for the developed hydro gazetteer in this thesis. Standards for geographic information representations developed by the International Organizations for Standardization (ISO) are used as a foundation. This chapter discusses the concepts and relationships borrowed from the ISO standards and used in the developed *GazOntology*.

3.1 Geographic Information in Gazetteers

Gazetteers are directories of features with geographic attributes and are sources of informal geo-referencing (Buchel & Hill, 2011). ISO recommends two standards for describing the spatial references in geographic information which relate a feature to the real world: (1) spatial referencing using coordinates (ISO, 2002) and (2) spatial referencing using geographic identifiers (ISO, 2003).

Gazetteers identify each feature with a location instance in a spatial reference system. The position of a feature is identified by this spatial reference which may be a GeoIdentifier or Coordinates. This spatial reference is stored as an attribute of the feature within a geographic dataset. The attribute used as a spatial reference uniquely identifies geographic information of the real world feature. These geographic attributes may be a set of location types or a hierarchy of location types making taxonomy based spatial querying possible or a set of coordinates allowing for spatial querying and analysis.

Geographic names of features, postal codes, or river basin names are examples of geoidentifiers that can be used to locate features. Harris County is an instance of a GeoIdentifier, which is the official name identifying a county in Houston. Spatial referencing by coordinates captures the spatial foot prints as points, lines, or polygons and the associated coordinate reference system in order to associate a physical space on the surface of the earth. Typically a feature may have multiple geometric representations in a gazetteer.

3.1.1 Spatial Referencing Using Geoidentifiers

Spatial referencing using geoidentifiers is based on the relation between a geographic feature and a location instance and may be descriptive in nature. The relation of the position to the feature may be a containment relation, where the position is described relative to a larger geographic feature for example a town contained in a state or it may be based on a relative measurement such as a given distance along a street from the cul-de-sac or fuzzy relations with geographic feature such as adjacent to a building (ISO, 2003). A spatial reference system using geoidentifiers is comprised of location types along with their geoidentifiers. These location types may be related to each other forming a hierarchy. Below is an example from the Getty Thesaurus of Geographic Names (TGN), which describes Houston. TGN uses one hierarchical containment relationship starting at *World* to represent places within political location types and physical location type hierarchies.

ID: 7013727   Record Type: administrative

Hierarchy of Houston (inhabited place)   Houston (inhabited place)

Coordinates:

Lat: 29 45 00 N degrees minutes   Lat: 29.7500 decimal degrees
Long: 095 21 00 W degrees minutes   Long: -95.3500 decimal degrees
Note: ***Connected with Gulf of Mexico by huge ship canal****; early center was destroyed by Mexican general Santa Ana in Texas Revolution; during American Civil War city was refuge for blockade escapees; completion of canal & discovery of oil stimulated growth.*

Names:

Houston (preferred,C,V) : named for Sam Houston (died 1863), American general, politician & president of the Republic of Texas
Harrisburg (H,V)
Houston City (C,V)

Hierarchical Position:

Hierarchy of World (facet)   World (facet)
Hierarchy of North and Central America (continent)   ....   North and Central America (continent) (P)
Hierarchy of United States (nation)   ....   United States (nation) (P)
Hierarchy of Texas (state)   ....   Texas (state) (P)
Hierarchy of Harris (county)   ....   Harris (county) (P)
Hierarchy of Houston (inhabited place)   ....   Houston (inhabited place) (P)

Place Types:

inhabited place (preferred, C)   ....   settled in 1824, expanded greatly in 20th cen.
city (C)   ....   incorporated in 1837
county seat (C)
port (C)   ....   ***now a deep water port,*** *was connected to Gulf of Mexico by*

3.1.2 Spatial Referencing Using Coordinates

Spatial referencing using coordinates are based on sets of X, Y, Z and M (linear measurement) coordinate values representing the positional geometry of a geographic feature in geographic space (ISO, 2002). Coordinates will unambiguously identify a physical geographic location in the map or on the earth, if a coordinate reference system is associated with the coordinates. Cartesian, Projected, Geographic, Geodetic, Polar, Horizontal and Vertical coordinate systems are some of the common Coordinate Reference Systems in use today. ISO standards define all the elements that are necessary to fully define a coordinate reference system associated with geographic information. By explicitly storing coordinate reference systems and the transformations needed to convert to other coordinate reference systems, aligning geographical data in different coordinate systems is possible for integrated search and analysis.

3.1.3 Location Equivalence in Spatial References

When using spatial references, various location types are possible for a given feature. For example, a river can be identified with an official name, alternate name, feature id, or river reach code. Also, a river may be represented as a polyline or polygon based on the nature of an application. Hydrologists may be interested in the length of the river reach or the area inundated by the river during flooding depending on their domain of analysis. Hence it is useful to represent multiple spatial footprints of geographical features and it is equally important to establish location equivalence between different possible spatial references of the same feature. Current digital gazetteers implicitly make this association. Figure 3.1 illustrates the concept of location equivalence between spatial reference types that locate the same feature.

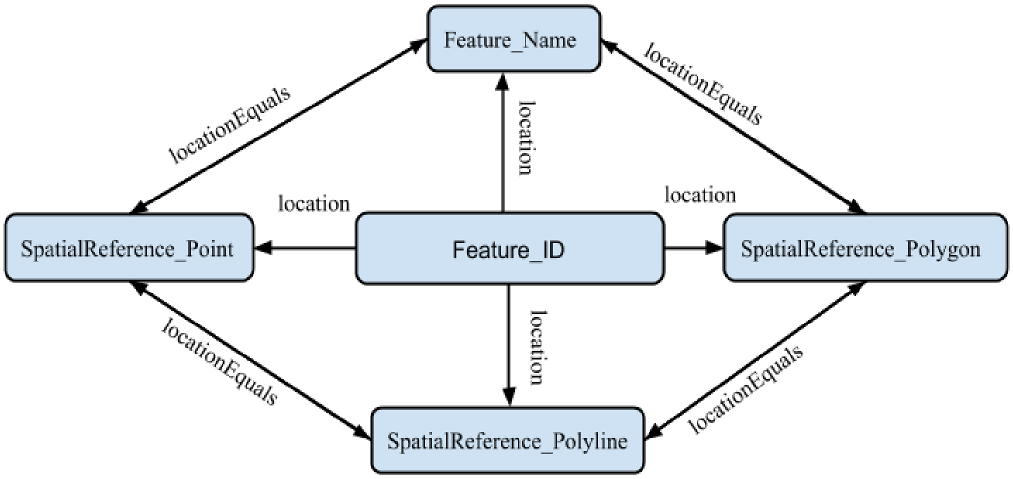


Figure 3.1 Location Equivalence in Spatial References

3.2 GazOntology

The GazOntology developed in this work, represents the gazetteer concepts and relationships. A feature is represented by a unique identifier for the feature. GeoIdentifier and Geometry are defined as specializations of a SpatialReference class. GeoIdentifier and Geometry subclasses are adapted from the ISO Standards for spatial referencing by geoidentifiers and spatial referencing by coordinate respectively. Features can have one OfficialName and zero to many AlternateName instances under specialization of the GeoIdentifier class. An OfficialName is the name officially recognized by a national naming authority, which in the US is the US Board on Geographic Names. AlternateNames can reflect local or historical variants of feature names. Figure 3.2 represents the class hierarchy of the GazOntology

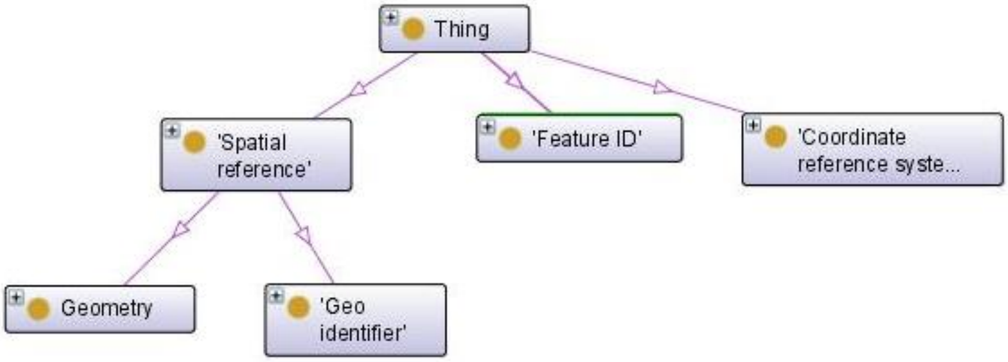


Figure 3.2 Class Hierarchy in GazOntology

The Geometry class represents the coordinates of the features as Well Known Text (WKT) literals and also stores the geometry type explicitly as point, polyline or polygon types. Table 3.1 shows example WKT representations for a point, line and polygon with x,y,z,m parameters

Feature Name	WKT Literal
Thirtyfoot Falls	POINT ZM (-69.105525603174442 46.412412994625356 0 NAN)
Snake Brook	MULTILINESTRING ZM ((-69.104614069842512 46.268929128181355 0 100, -69.101971003179926 46.268695128181719 0 87.236969999999999, -69.10188720318007 46.268653594848445 0 86.743750000000006))
Norway Pond	MULTIPOLYGON ZM (((-68.98401507002967 46.428501994600367 0 NAN, -68.98401460336305 46.428273394600751 0 NAN, -68.983716803363507 46.428090794601019 0 NAN, -68.983319803364111 46.427976927934537 0 NAN..)))

Table 3.1 WKT Representation of Feature Geometry

The CoordinateReferenceSystem class contains basic coordinate reference system information such as well-known id and well known text representation of the coordinate system parameters as string literals. A coordinate system contains keywords for coordinate type (PROJCS for projected coordinates, GEOGCS for geographic coordinates, or GEOCCS for geocentric coordinates). The keyword is followed by terms that define the coordinate system. For a projected coordinate system this includes a projection name followed by the geographic coordinate system, the map projection, one or more parameters, and the linear unit of measure. The WKT expression for NAD83 is shown below.

NAD_83	GEOGCS["GCS_North_American_1983",DATUM["D_North_American_1983",SPHEROID["GRS_1980",6378137.0,298.257222101]],PRIMEM["Greenwich",0.0],UNIT["Degree",0.0174532925199433],AUTHORITY["EPSG",4269]]

Table 3.2 WKT Expression for NAD83 Coordinate System

Well known text literals are chosen as the desired format to store spatial information in order to ensure that the developed triple store is compatible for querying in GEOSPARQL in the future.

The GazOntology has a *locationEquals* property which establishes location equivalence between different spatial reference types. The *locationEquals* property is modelled as a symmetric and transitive OWL property with domain and range as SpatialReference and its subclasses. This property ensures that two instances of spatial reference class will identify the same geographic feature and each of these location representations can be used appropriately based on the context of search. Figure 3.3. presents a UML model of the GazOntology classes and relationships.

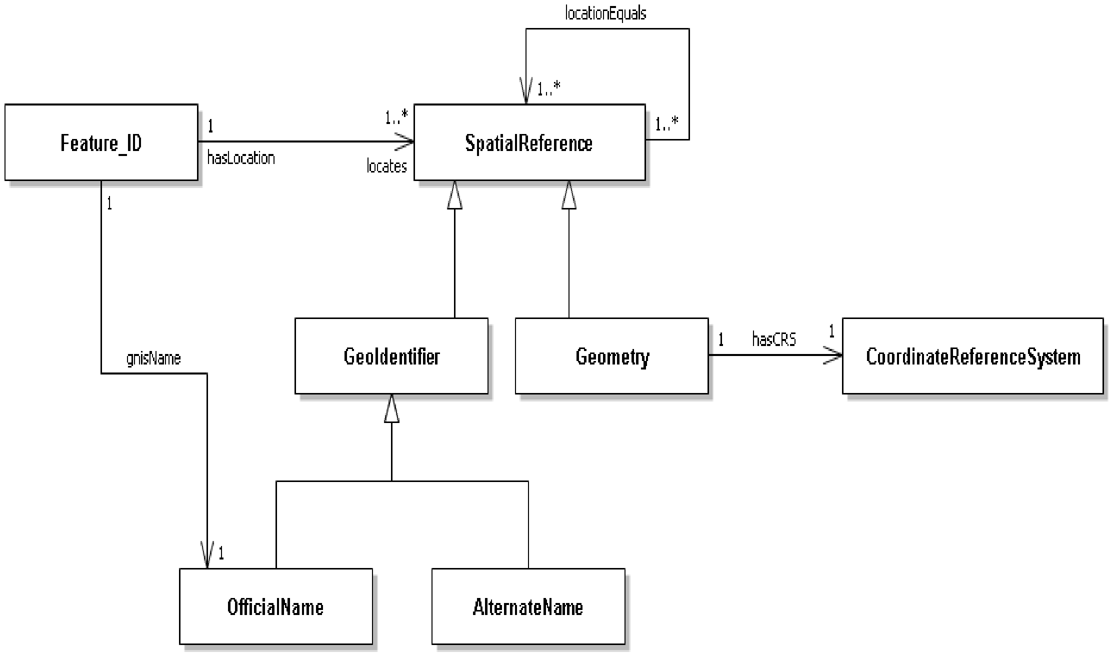


Figure 3.3 UML Diagram of GazOntology

CHAPTER 4

HYDROLOGICAL FEATURES ONTOLOGY

The gazetteer ontology described in Chapter 3 models digital representations of geographic features through their descriptors, which include their unique identifiers and spatial references, which include their official and alternate names and coordinate representations. The goal in developing a semantically enhanced gazetteer is to add richer semantics based on feature types so that the gazetteer is able to support queries on features and between features based on semantically appropriate relationships. The spatial relations defined in RCC8 and through the 9-intersection represent pairwise relations between geometry types (e.g., region-region, line-line, line-region). The unit of representation in a gazetteer is a named feature so the interest of this thesis is to identify semantic relationships among features that are pertinent to the feature types. This chapter considers relationships among feature types as they exist in the world and as they may be expressed by natural language. The following chapter then considers how such relationship map onto the well-defined 9-intersection model relationships. The goal of this chapter is to provide a domain model of canonical geographic feature types and relationships between them. For scoping purposes, the chapter considers the domain of surface hydrology, and canonical feature types of this domain. The core of this chapter is the specification of canonical surface water feature types and identification of relationship among these feature types. This chapter begins by providing a context for the developed ontology within other existing ontologies. The ontology was designed to align with upper level ontologies such as DOLCE and Basic Formal Ontology. The chapter also describes how the ontology aligns with a recent ontology design pattern developed for surface water features (Sinha, et al., 2014) and outlined in Chapter 2. The chapter proceeds to identify the canonical forms of surface hydrological features and elaborates on hydrological relationships between these features. This chapter then describes the Hydrological Feature Ontology concepts and relationships.

4.1 DOLCE Upper level Ontology

DOLCE is the first module of a Library of Foundational Ontologies being developed within the Wonder Web project. DOLCE aims at capturing natural language underlying the ontologies combined with human common sense (Gangemi, Guarino, Masolo, Oltramari, & Schneider, 2002). DOLCE is considered the starting point to model the hydrological features, relationships and their meanings as a domain ontology.

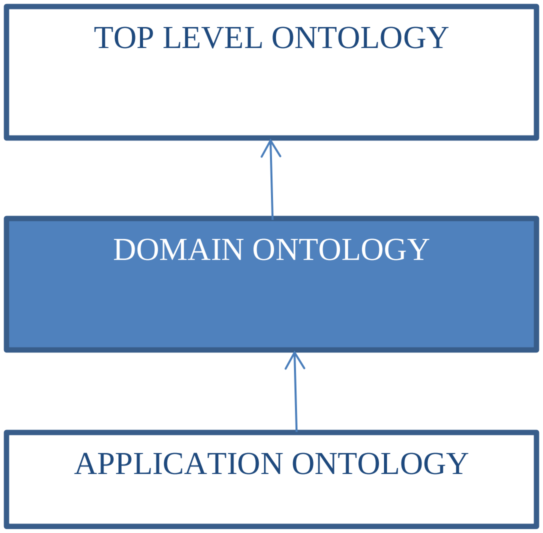


Figure 4.1 Three Level Architecture of Ontology

DOLCE, like other upper level ontologies, makes a fundamental distinction between endurants and perdurants. Endurants are wholly present at any given time along with their parts. Perdurants are entities that extend in time with multiple temporal parts and at any given time, only some of their temporal parts are present. Participation is the primary relationship between endurants and perdurants. The surface hydrology domain ontology links to the DOLCE top level categories as represented in Figure 4.2.

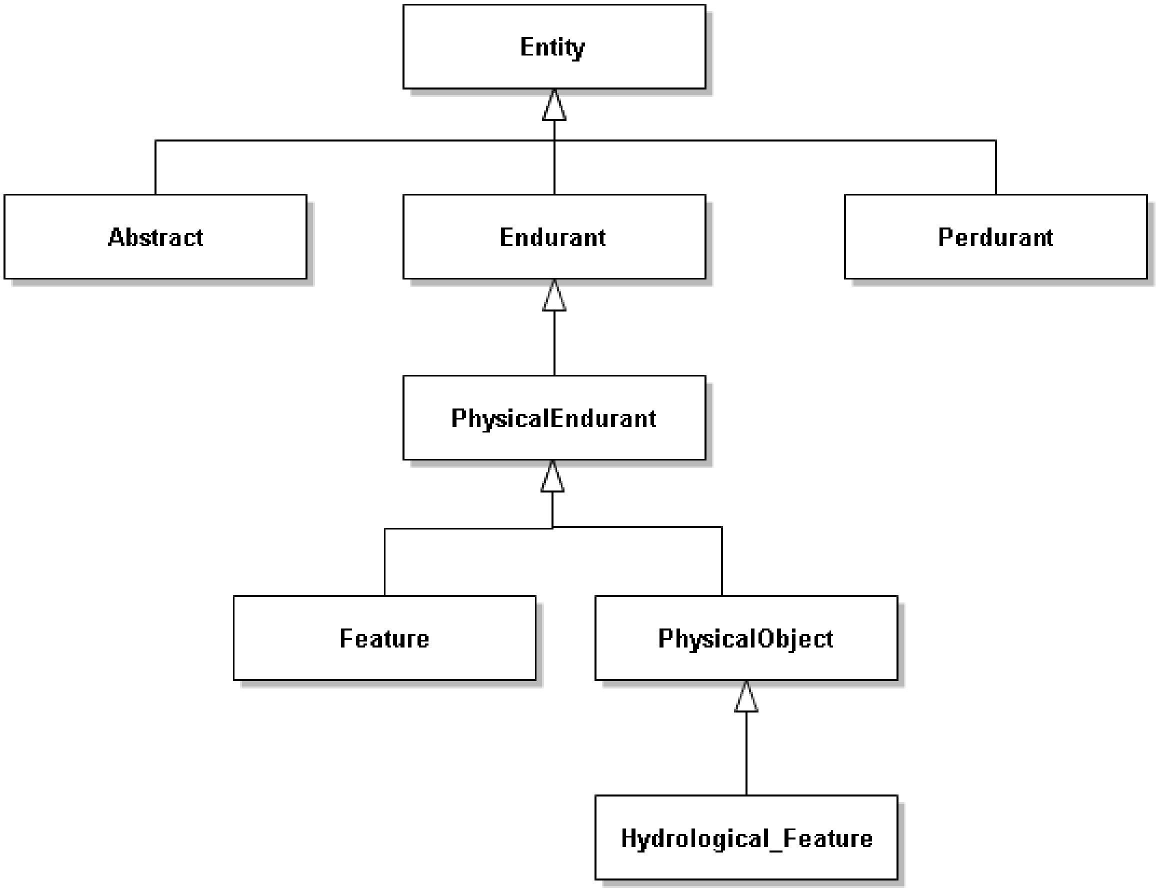


Figure 4.2 HydoOntology Links to DOLCE Classes

DOLCE defines *PhysicalObject and Feature* as subclasses of *PhysicalEndurant* since they have direct spatial properties. A Feature in DOLCE however, has a different connotation than feature as understood in this thesis. In DOCLE a feature is considered as (Gangemi, Guarino, Masolo, Oltramari, & Schneider, 2002) a “parasitic entity”, such as a hole or bump in a road. Such features are considered wholes, but no common unity criterion exists for them (Gangemi, Guarino, Masolo, Oltramari, & Schneider, 2002). *PhysicalObjects* are endurants with unity where unity refers to a property that uniquely identifies the parts of an instance. Different *PhysicalObjects* may have different unity criteria. Further each *PhysicalObject* does not depend on other *physicalobjects* for their existence (Devaraju & Kuhn, 2010). Typical examples are Waterbody, Riverbasin, and River This thesis thus identifies and models the surface hydrological features as a subclass of *PhysicalObjects.*

4.2 Surface Hydrological Features

Surface hydrology deals with the flow of water and its constituents over the land surface (Chow, et al., 1988). This branch of hydrology is concerned with surface runoff in streams, lakes, rivers, and reservoirs. Modelling such spatially dynamic features and processes has been accomplished by incorporating the dimension of time and representing the change factor in geographic information databases (Goodchild, 2000). The surface hydrology ontology design pattern (Sinha, et al., 2014) builds a foundation for such dynamic behaviour of flowing surface water. As discussed in Section 2.4, the canonical forms of hydrological features characterized in the following sections can be seen as specializations of the Wet Semantics module. The Wet Module of the surface hydrology design pattern makes a fundamental separation between flowing and non-flowing feature types, a separation utilized in the developed ontology. The gazetteer context of the thesis provides an additional domain focus for the ontology development in that the units of interest are named features or feature parts and natural language expression for relations between such features.

4.2.1 General Hydrologic Network Components

General hydrological network components describe flow relations, typically starting from some source. A source is where surplus water enters the surface water system, which is usually a stream, river, or a catchment. A source is considered to be the origin of a water body and participates in the run off processes, by contributing to surface runoff caused by precipitation.

A sink or mouth is where the surplus water leaves the surface water system. This is the point where a water body discharges into another water body or infiltrates into the subsurface contributing to the groundwater table or aquifer as subsurface runoff.

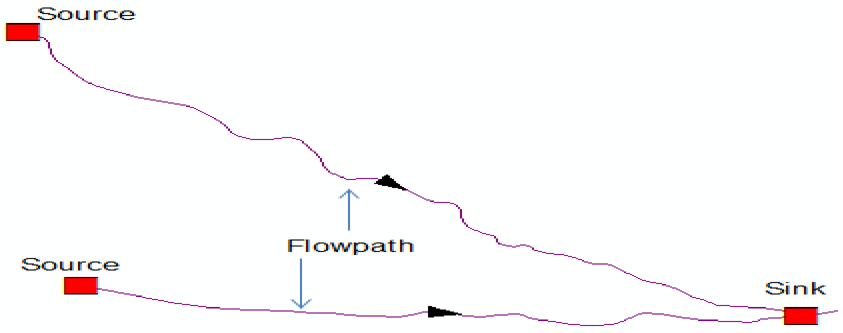


Figure 4.3 Source, Sink, and Flowpath

A general principle articulated in the surface hydrology design pattern (Sinha, et al., 2014)is a distinction between the terrain surface structures that act as containers and channels and the water that is resident or flowing through them. Primary classes in Sinha et al. include Channel, Depression and Interface in the Dry Module and associated Stream segment, Waterbody, and Fluence in the wet module. A channel is a linear feature that accommodates flow from source to sink creating a flow path and thereby establishing a flow direction. A junction is the point of confluence of two or more channels or the bifurcations of one to more channels. Such junctions are represented by the Interface and Fluence classes in Sinha et al. These transitional concepts may be useful for detailed hydrological modelling but do not support identification of direct relationships between named features as desired for gazetteer queries. The HydroOntology developed in this thesis captures relationships between feature types directly through OWL properties.

Source, sink, and junction for example are conceptualized as OWL properties connecting surface hydrology features and allowing multiple hydrological features to be specified as domain and range values. Flow relations discussed in the section below are similarly conceptualized as OWL properties connecting feature types.

4.2.1.1 Rivers and Streams

Streams and rivers are large persistent channelized flows of water into other waterbodies. Because they share the common behaviour of channelized flow they constitute a canonical feature type which will be referred to by the term Stream. Streams typically have headwaters, which are the source for the stream, a main stem, zero or more tributaries which are smaller streams and a mouth representing a sink that may be an ocean, lake, or another larger river.

The beginning of the river is known as its headwaters and it is the source for the river. Sources for the rivers may be a spring that is fed by ground water, a glacier with melting ice or a set of streams that converge into a water body. More often the headwaters of a large and powerful river are a small pond or a trickling stream.

A river network generally resembles a tree structure, with a *main stem* and many tributaries that bifurcate further into smaller rivers and streams. A *tributary* is a river or stream that *flows into* another river. Large rivers are fed by many tributaries. The point where two or more channels *merge*, for example where a tributary joins a main river, is called a confluence and when a channel *diverges* into two or more streams it is called divergence. Points of confluence and divergence are often called junctions as they connect two or more channels.

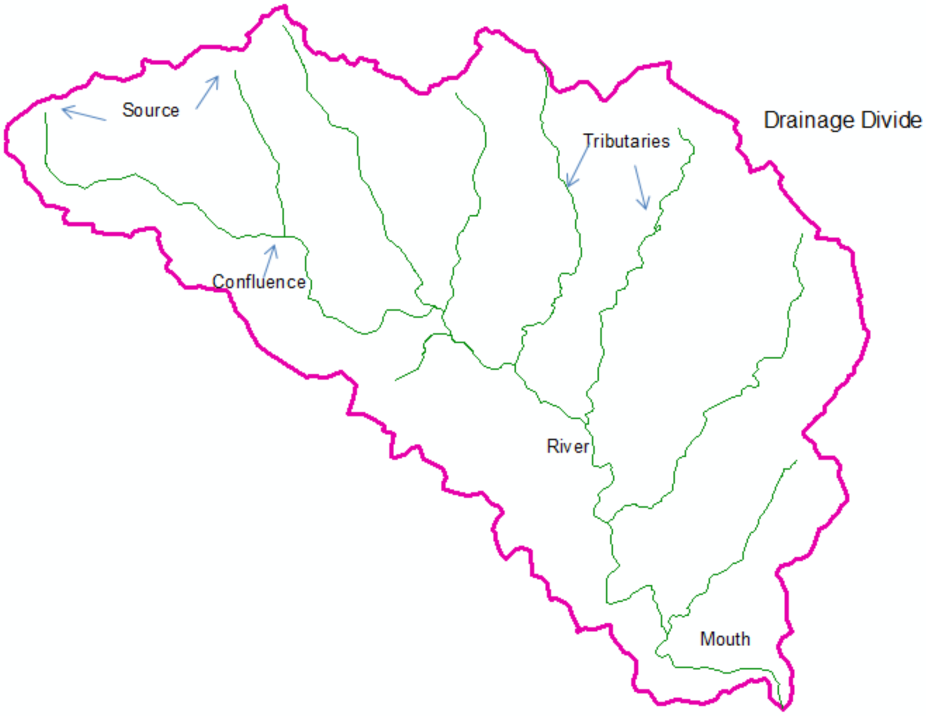


Figure 4.4 Parts of a Stream Network

Stream order and stream length are important spatial properties used to model the streams within a network. Streams that originate at the source are called first order streams. When two streams of order n merge, the resulting stream is of order (n +1). When streams of different order merge, the resulting downstream is of order (max (n1, n2)).

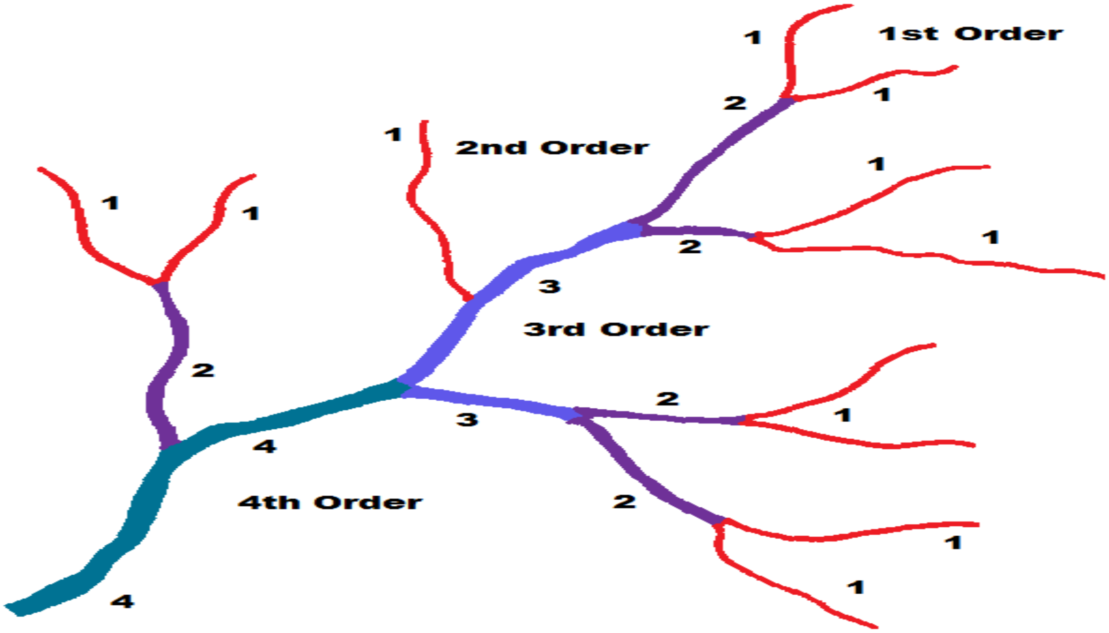


Figure 4.5 Stream Order

The Mouth of a stream is where it empties into another water body. While a mouth can be considered a transition point it is common to refer to a mouth in terms of the named receiving body of water. A mouth may thus be another river, a lake, an ocean or a named part of an ocean. The Gulf of Mexico, for example, is typically named as the mouth of the Mississippi River.

4.2.1.2 Lakes, Ponds and Reservoirs

Lakes and ponds are water bodies formed by storing run off water in depressions on the land surface. Lakes are usually surrounded by land and are fed and drained by rivers and streams. In contrast with the rivers which are flowing bodies of water, lakes have very slow water velocities. Lakes are typically located along the course of a river system and have a drainage basin, inflow and outflow. Ponds are typically smaller bodies of water than lakes but otherwise share the same hydrological behaviors of lakes. Reservoirs are also contained water bodies but constructed by humans rather than naturally occurring and their inflow and outflow may be regulated by dams and weirs.

4.2.1.3 Springs

Springs are naturally occurring discharge features of groundwater flow systems. Groundwater flow to springs (and therefore the characteristics of the source area) is governed mainly by three inter-related factors: geology, topography (landforms and relief), and climate (timing and amount of precipitation) by influencing the amount of water that occurs as surface flow versus the amount that infiltrates into the ground as recharge to groundwater. All three factors govern how the subsurface flow system develops and where springs occur.

4.2.1.4 Wetland

Swamps and marshes are wetlands that are usually found in flood plains. Wetlands are the transitional zone between aquatic features and terrestrial ecosystems and are usually saturated with water harboring a habitat on its own. Wetlands also store flood waters during flooding seasons and are fertile grounds due to sedimentation. They are fundamental hydrologic landscape units (Winter, 2001) and can be defined topographically as landform with flat areas or shallow slopes that lie *adjacent* to perennial water bodies and are inundated by these water bodies.

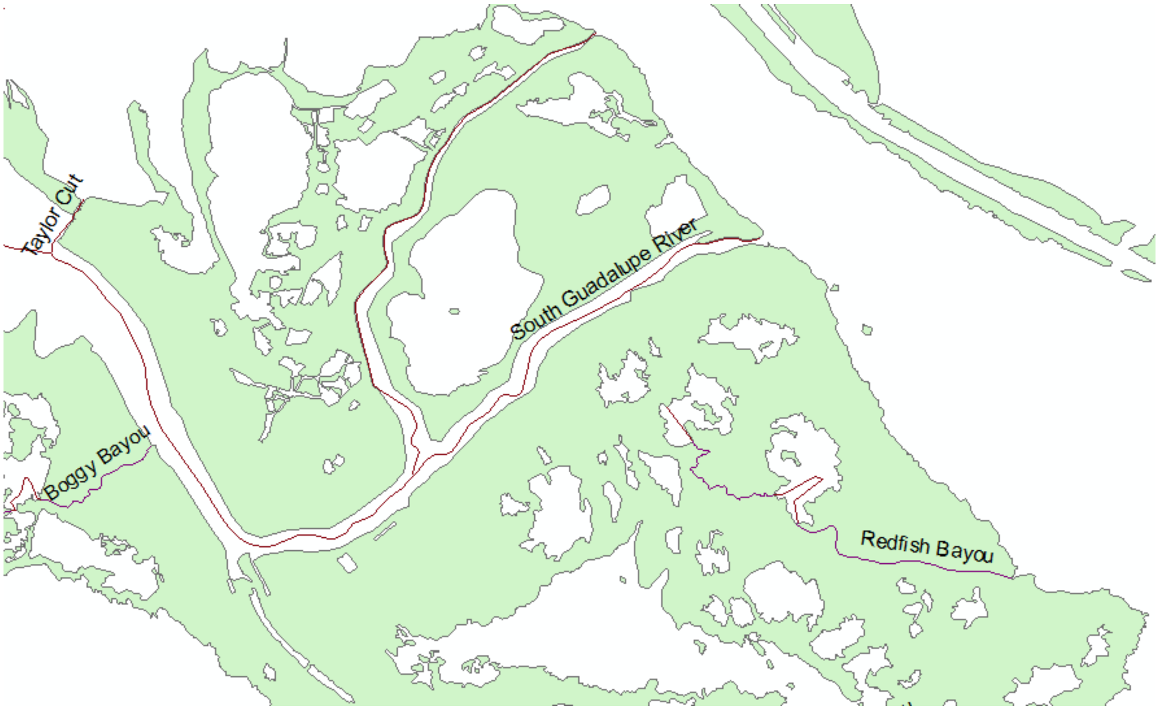


Figure 4.6 South Guadalupe River with Wetlands

4.2.1.5 Coastal Features

Coastal landforms are valuable environmental resources and play an important role in recreational and maritime activities. The coastal zone is a very dynamic environment, where the land surface is constantly subjected to wave action and ocean currents. The combined effect of waves, currents, and tides causes various geomorphological processes to alter the size and shape of the coastal zone. Abrasion is the most dominant process, caused by the scraping or impact of sediment carried by water against the shore. Coastal landforms can be categorized based on the processes that create them as depositional landforms and erosional landforms. The waters enclosed by these various land forms are associated with a number of feature types.

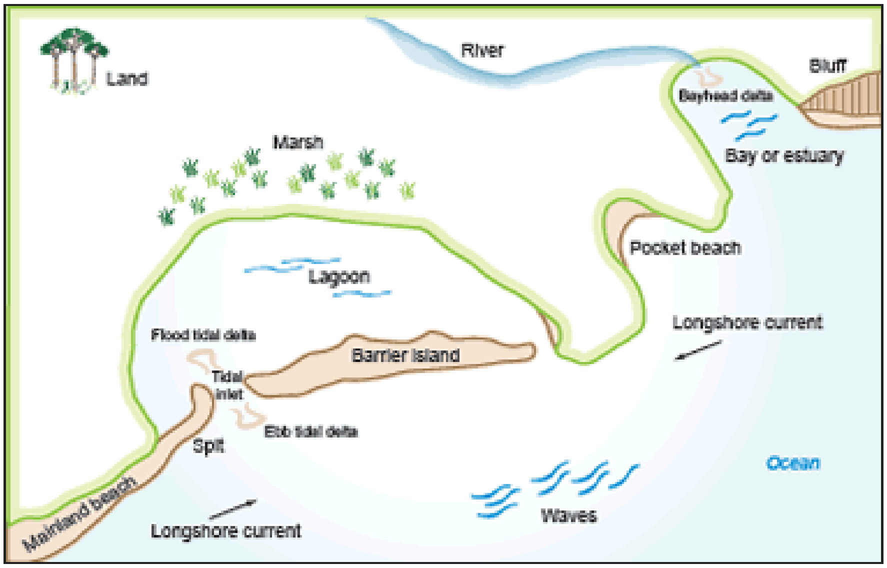


Figure 4.7 Coastal Features

A bay is large body of water which is a part of the ocean or sea formed by a shoreline indentation. A larger bay is called a Gulf, cove or sound and if a bay exists within a gulf, a bay can be modelled as a part of the gulf. For example the Gulf of Mexico contains Galveston Bay area as well. If a bay is separated from the ocean, by barrier islands, then the formation is called a lagoon (Figure 4.7). A harbor is a part of the ocean or sea closer to land and deep enough for ships, vessels, boats and barges to be docked safely. Harbors can be natural or artificial. Ports are usually located in harbors for loading and unloading vessels. All of these features share the common behavior of being parts of an ocean and potentially in hierarchically nested relationships.

Similar to coastal features which have evolved into variously named parts of an ocean, freshwater bodies may also have named parts. Large lakes may include bays or coves with similar definitions to their coastal counterparts. Rivers have also evolved named parts that can include bends and elbows.

4.3 HydroOntology

Based on these prototypical surface hydrology feature types and parts, a HydroOntology was developed as an OWL domain ontology. These real-world surface hydrological feature types are modelled as OWL classes. Subsumption relations are realized by sub classing entities through the RDFS subclass relationship and part hood relations expressed as OWL object properties and sub properties. The focus of this thesis is to semantically enable a gazetteer of hydrological features for place name searches; hence the feature classes captured in this ontology are basic components of a hydrological network which may have feature names and hence be modelled in a gazetteer.

The HydroOntology captures real-world hydrologic features and topological relations between features. By capturing topological relations both connectivity relations and flow relations in a hydrologic network can be modelled. Hydrographic_Feature, and Boundary form the top level classes of the ontology. Hydrographic_Feature contains two disjoint classes FreshWater_feature and SaltWater_feature. Boundary class has Coastline and Watershed classes which are named features. All the features that are identified and described in section 4.2 are modelled as specializations of the top level Hydrographic Feature class.

Basic ontological relations can be realized between endurants modeled in the developed HydroOntology. The term individual or particular refers to entities that cannot have any instances and the term universal refers to entities which can have instances. Instantiation relations exist between particulars and universals. All the named hydrological features populated in the resultant hydro-gazetteer are instances of HydroOntology classes. For example, Machias River (particular) is an instance of River class (universal). Pushaw Lake is an instance of Lake class.

Subsumption relation between two universals or classes implies that all individuals of one universal are individuals of the other universal. For example, all instances of River class are necessarily instances of FreshWater_feature class as well as FreshWater_Feature subsumes the universal River.

The class hierarchy of the HydroOntology is shown in Figure 4.8. Two kinds of parthood relations are distinguished in DOLCE which hold among concrete entities. Temporary Parthood exists between two endurants where one endurant is a part of another endurant. For example, Rapids isPartOf River, River hasPart Falls. Temporal parthood is a time dependent relation that exists between perdurants. For example, Precipitation is PartOf Rain.

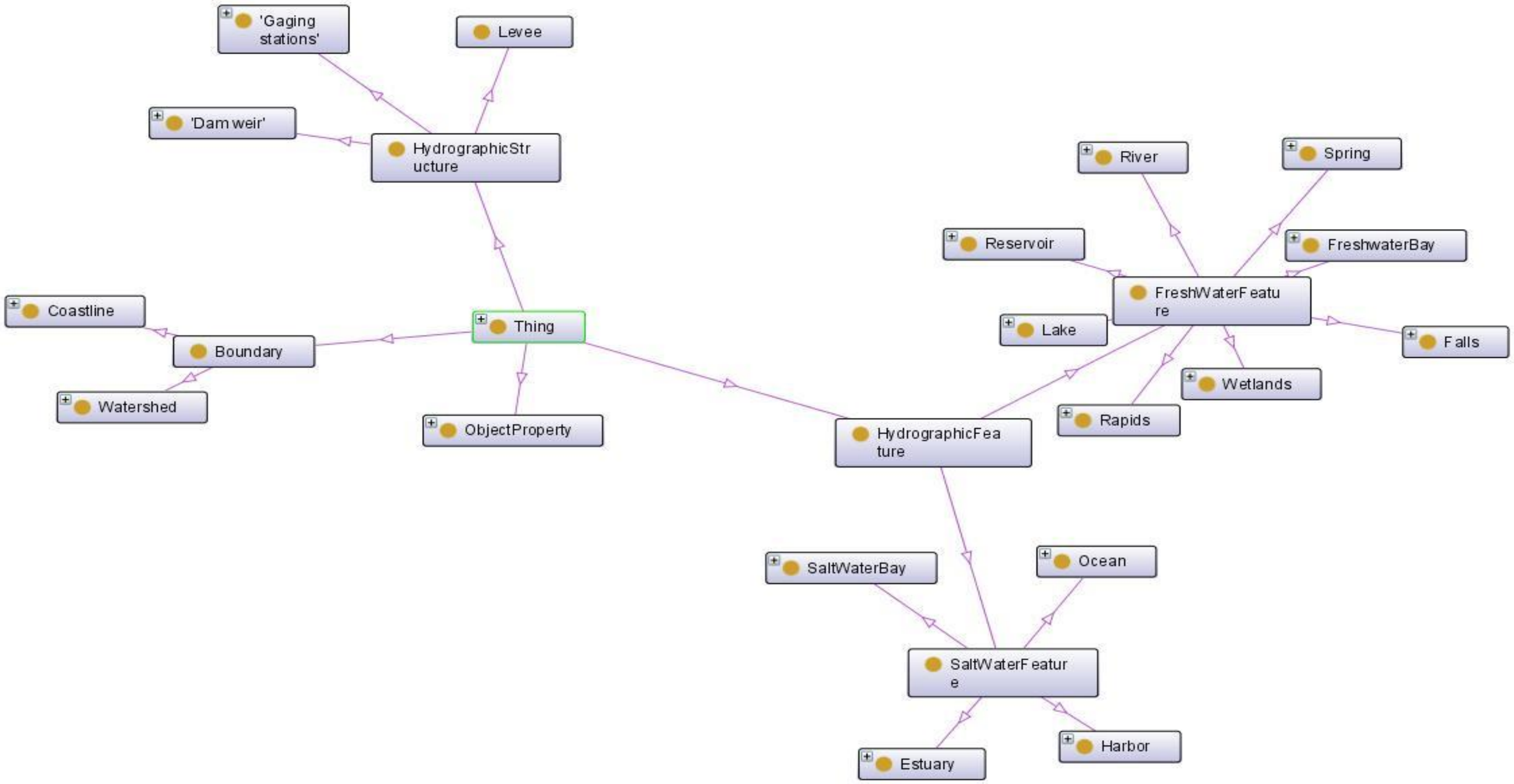


Figure 4.8 Class Hierarchy of HydroOntology

A river or stream has a source or headwaters, a mouth which is a discharge point, and zero to many tributaries. These relations express that a river and other hydrologic features are geographically connected and contribute inflow or outflow of water to each other. Sources of a river can be other hydrological features such as a Spring, River or Lake. Similarly the terminal feature of a river can be another river, a lake, a Bay or an Ocean. By specifying object properties between feature classes, such as *River hasSource Spring* , *River hasMouth Lake* , *River hasTributary River,* it becomes possible to query and make inferences from the hydro-gazetteer on hydrological relationship between proper named features.

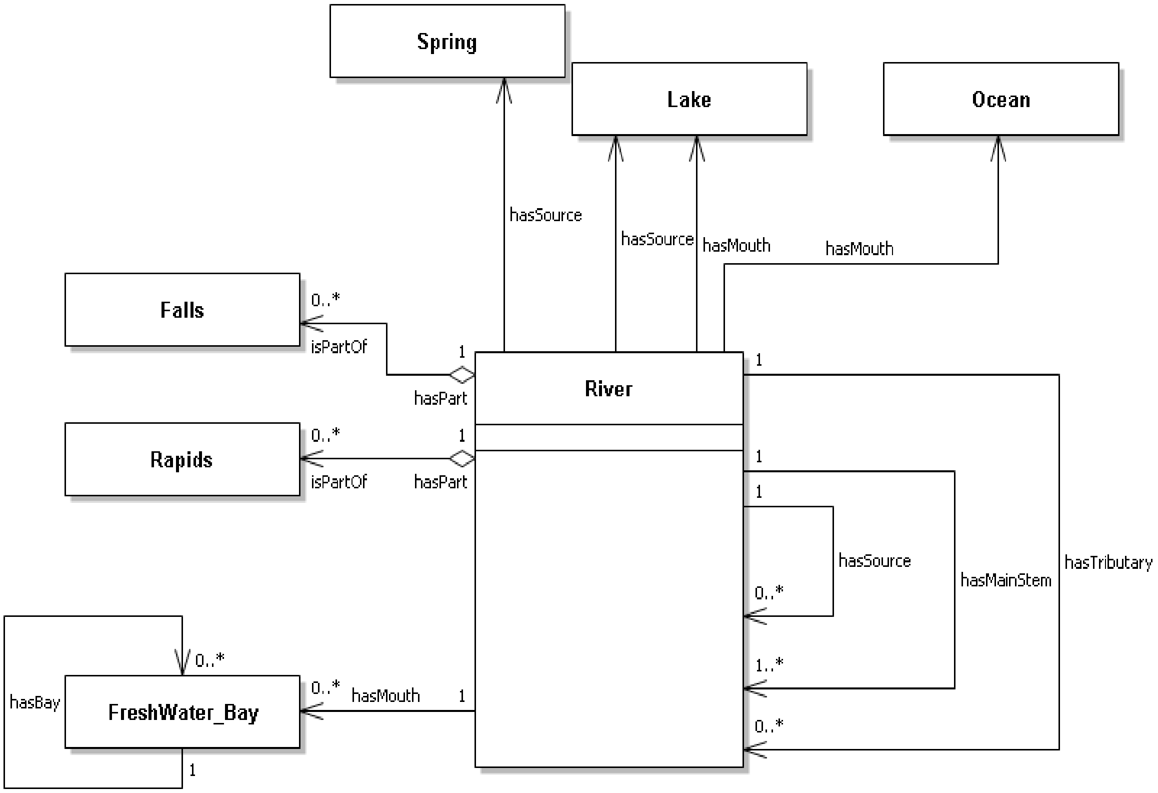
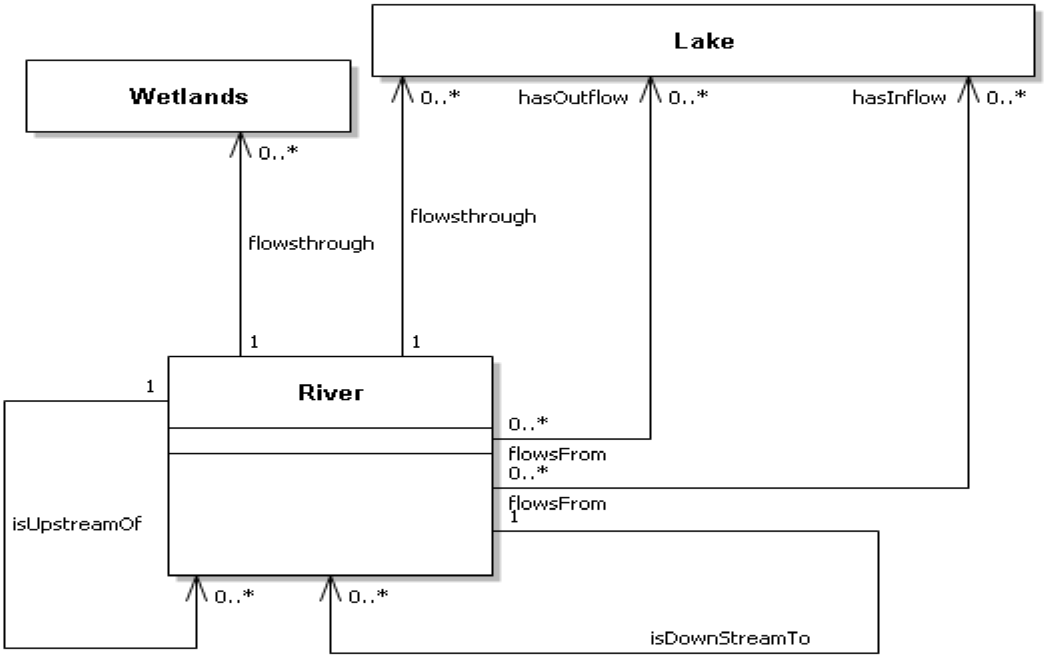


Figure 4.9 Hydrographic Relationships of River to Other Feature Types

A number of flow relations can be modelled between River, Lake and Wetlands classes. A river *flowsthrough* Wetlands or Lake. A lake can have *inflows* and *outflows* which are rivers and inversely River *flowsinto* and *flowsfrom* Lake. Flow navigation can also be modelled as transitive properties *isUpstreamOf* and *isDownstreamTo.*



.

Figure 4.10 Hydrographic Relations of River

Bays, especially Saltwater_bays, can contain zero or many bays, coves, or fjords. Hence transitive properties and corresponding inverse properties *hasSaltWaterBay /isSaltWaterBayOf* and *hasFreshWaterBay / isFreshWaterBayOf* can model nested bays.

Object properties along with domains and ranges for each of the properties as modelled in the hydro-ontology are listed in the table below.

Property	Sub Property	Domain	Range
hasHydrologicPart	hasSaltWaterBay	SaltwaterBay,Ocean	SaltwaterBay
	hasFreshWaterBay	River, Lake	FreshwaterBay
	hasMainStem	River	River
isHydrologicPartOf	isFreshWaterBayOf	FreshwaterBay	River, Lake
	isSaltWaterBayOf	SaltwaterBay	SaltwaterBay,Ocean
	isMainStemOf	River	River
isHydrologicallyConnectedTo	isMouthOf	Bay, Ocean, Lake, River	River
	isSourceOf	Springs, River, Lake	River, Lake
	hasSource	River, Lake	Springs, River, Lake
	hasMouth	River	River, Lake, Bay, Ocean,
	hasTributary	River	River
	isTributaryOf	River	River
hasHydrographicRelation	flowsFrom	River	River, Lake
	flowsInto	River	River, Lake
	flowsThrough	River	Wetlands, Lake
	hasInflow	Lake	River
	hasOutflow	Lake	River
	isDownsteamTo	River	River
	isUpstreamOf	River	River
hasHydrographicStructure	hasDam	River	Dam
	isDamOf	Dam	River
	hasGagingStation	River	GagingStation
	isGagingStationOf	GagingStation	River

Table 4.1 OWL Properties with Domain and Range

4.4 Summary

This chapter presented a domain model of common hydrologic features and relationships between them. HydroOntology developed in this thesis, modelled these features as OWL classes and their relationships were represented as OWL properties and sub-properties. Class Hierarchy of the HydroOntology and OWL properties with their domain and range values were presented in detail. This chapter also showed, how hydrographic relations and flow relations between feature types can be modelled in an information system. Further, this chapter described how the developed Ontology aligns with other top-level ontologies such as DOLCE and with a recent ontology design pattern developed for surface hydrological features.

CHAPTER 5

SPATIAL RELATIONS IN SURFACE HYDROLOGY

The surface hydrology ontology described in Chapter 4 identified relations between prototypical surface hydrology features classes. The interactions between these hydrological features that result in the exchange of surface water can be seen as specializations of topological relations. By modelling semantics of topology relations in a dynamic environment like hydrology, geographic information retrieval can be expanded to cover relationships between named features.

In order to determine what instances of surface hydrology features participate in these relations, information is obtained through spatial analysis of geographic data sets that include the National Hydrography Dataset, the NHDPlus data set and the Geographic Names Information System. The 9^+-intersection model (Kurata & Egenhofer, 2006) (Kurata & Egenhofer, 2007) between point, line, directed line, and region geometries identifies sets of relationship between geometries independent of what these geometries represent in the world. This chapter considers the semantics of these relations in a hydrological context as presented in HydroOntology and explains the spatial analysis methods used to extract these relations from geographic data sets. The chapter starts with a brief description of the supporting geographic data sets.

5.1 Geographic Names Information System

USGS along with US Board on Geographic Names maintains the Geographic Names Information System (GNIS) which is the primary source of identifying official place names. GNIS is the Federal standard for geographic nomenclature. There are over 2.5

million names currently in the database (http://geonames.usgs.gov/docs/pro_pol_pro.pdf). These records include names of natural features, populated places, civil divisions, mines, churches schools, dams, airports, and shopping centers except roads and highways. The GNIS feature ID is the only standard Federal key for identifying a specific geographic feature and the GNIS Feature Name and spatial footprints are the official feature attributes for federal use. The National Map Gazetteer is a geographic dictionary for domestic features that allows users to query based on feature types, state and county using information in the Geographic Names Information System. The GNIS_ID and GNIS names for hydrographic features are included in the NHD as described below.

The GNIS database is searchable for features using fundamental attributes, such as feature names, variant names, identifiers, state, and feature types. Features are also searchable based on the 1:24,000 scale USGS topographic map name, where the feature is located. Features are stored as point geometries and can be downloaded along with non-spatial attributes as pipe delimited files. All the coordinates are in North American Datum 1983. The downloaded features can be imported into GIS software and form one of the data sources for spatial analysis.

5.2 NHD Data Model

The National Hydrography Dataset is a vector geospatial thematic layer that represents the surface water component of the National Map. It is available nationwide as medium resolution at 1:100,000 scale, as high resolution at 1:24,000 and is also becoming available in selected areas on larger scales such as 1:5,000-scale mapping (http://pubs.usgs.gov/fs/2009/3054/pdf/FS2009-3054.pdf). The NHD Dataset is created by the cumulative effort of federal, state, and local government agencies. This partnership has resulted in a common data model, pooling of resources, and improved data interoperability.

The NHD Geodatabase contains two feature datasets: (1) Hydrography Dataset which contains point, line, and polygon feature classes to represent hydrological features and associated relationship classes to metadata. The Hydrography dataset also contains Point, Line and Area Events Feature Class which are not populated at this time. (2) Watershed Boundary dataset, which contains Hydrologic Unit Features.

A subset of the schema, which is common to most feature classes, is presented first, so that important attributes that are used for spatial analysis later in this thesis can be introduced.

Field Name	Description
Permanent Identifier	Identifier of the NHD feature. This field may also contain identifiers as 36 character strings in registry format formally known as Guids.
GNIS_ID	Unique identifier assigned by GNIS
GNIS_Name	Official feature name from GNIS
ReachCode	Unique identifier composed of two parts. The first eight digits identify the sub-basin code as defined by FIPS 103. The next six digits are randomly assigned, sequential numbers that are unique within a sub-basin, length 14.
FType	NHD Feature type
FCode	Numeric codes for various feature attributes in the NHDFCode lookup table

Table 5.1 Subset of Field Names Common for Hydrography Dataset

The Identifier used in NHD is a 10-digit integer value that uniquely identifies the occurrence of each NHD feature. Each value is assigned only once to a feature and once assigned, this value is associated permanently with that feature. If the feature is modified or deleted, the associated identifier is retired. Permanent Identifiers, if stored as registry style strings with 36 characters enclosed in curly brackets, are used to uniquely identify a feature or a record within a geodatabase and across geodatabases. GNIS_ID and GNIS_Name are populated from the Geographic Names Information System.

Reach Codes are also unique identifiers for a given feature, however they serve a different purpose than identifying features for spatial analysis or spatial data management in geodatabases. A reach is a continuous, unbroken stretch or expanse of surface water. In the NHD, a reach is defined as a segment of water surface that has similar hydrologic characteristics, such as a stretch of stream or river between two confluences, or a lake/pond (http://nhd.usgs.gov/chapter1/chp1_data_users_guide.pdf). Reach codes facilitate geocoding or linking observations and events to reaches. A reach code uniquely identifies each reach. This 14-digit code has two parts: (1) the first 8 digits are the hydrologic unit code for the sub basin in which the reach exists and (2) the last 6 digits are a sequence number assigned in arbitrary order to the reaches within that sub basin.

5.2.1 Hydrography Dataset

The hydrography feature dataset contains all the surface water feature classes along with a geometric network. Surface water feature classes are represented as point, polyline and area features and the geometric network represents the flow network.

5.2.1.1 NHDFlowline

The NHDFlowline feature class represents the complete linear flow network of the surface water drainage system and is the most important dataset that establishes flow relationships. Each record in the table represents a reach which is a stream segment between two confluence points with the same hydrological characteristics. The record includes line geometry, attributes to establish the flow direction and upstream/downstream flow relationships, and linear referencing measures to associate events at specific locations within the flow network. Table 5.2 shows the schema of NHDFlowline feature class.

Field Name	Description
Permanent Identifier /	Identifier of the NHD feature. This field may also contain identifiers as 36 character strings in registry format formally known as Guids.
LengthKM	Feature length in kilometres
FlowDir	Direction of flow relative to coordinate order. Values may be 'With Digitized' for known flow direction and 'UnInitialized' for unknown flow direction.
WBArea_PermanentIdentifier	Identifier of the NHD polygonal water feature through which an NHD "Artificial Path" flowline flows
FType	NHD Feature type
FCode	Numeric codes for various feature attributes in the NHDFCode lookup table
Shape_Length	Feature length in units of the spatial reference system
Enabled	Created when Geometric Network is built All features should be set to True (From the database).

Table 5.2 NHD Flowline Schema

The NHDFlowline feature types include Stream: a flowing body of water which may be intermittent, perennial or ephemeral, Artificial Path: a surrogate NHDFlow line feature to represent the flow of a named stream through a water body, Connectors: a known, but invisible connection of two non-adjacent network components, CanalDitch: An artificial waterway to connect two water bodies for irrigation purposes or for navigation, Coastlines: A line of contact between the open sea and the land, including imaginary lines separating inland water bodies from the open sea, Pipelines: A closed conduit, with pumps, valves and control devices, for conveying fluids, gases, or finely divided solids.

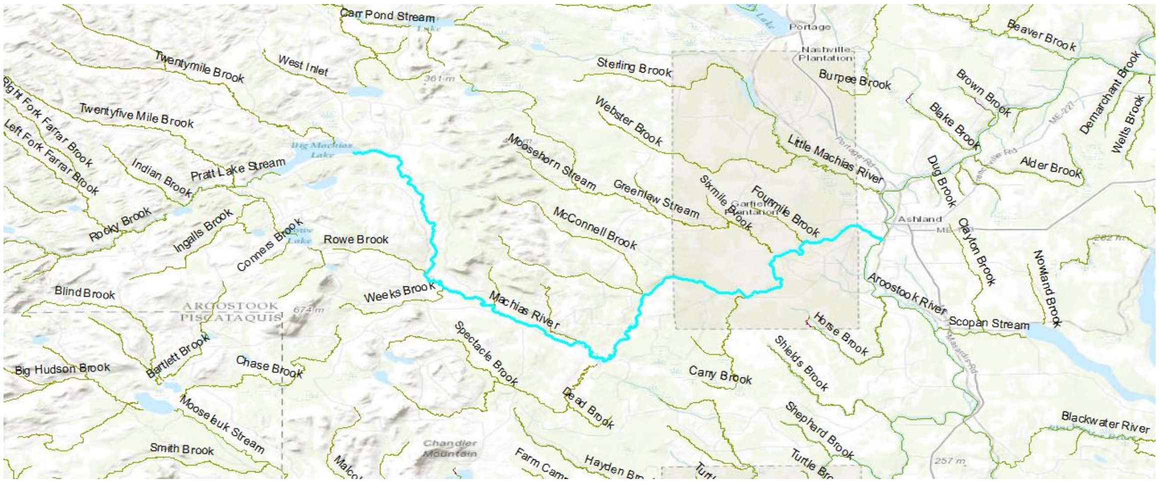


Figure 5.1 Machias River represented as a NHDFlowline

The main feature types that are of interest in this thesis are streams, artificial path, canals and coastlines. By representing streams as 1-dimensional line geometry, a hydrological network can be built, enabling network analysis to identify upstream and downstream segments and tracing flow path between two given points. In reality, streams occupy spatial extent and hence the NHD database includes multiple spatial representations of hydrologic features as a function of scale. NHDFlowline feature types stream and artificial path may have associated polygonal representations in the NHDArea feature class.

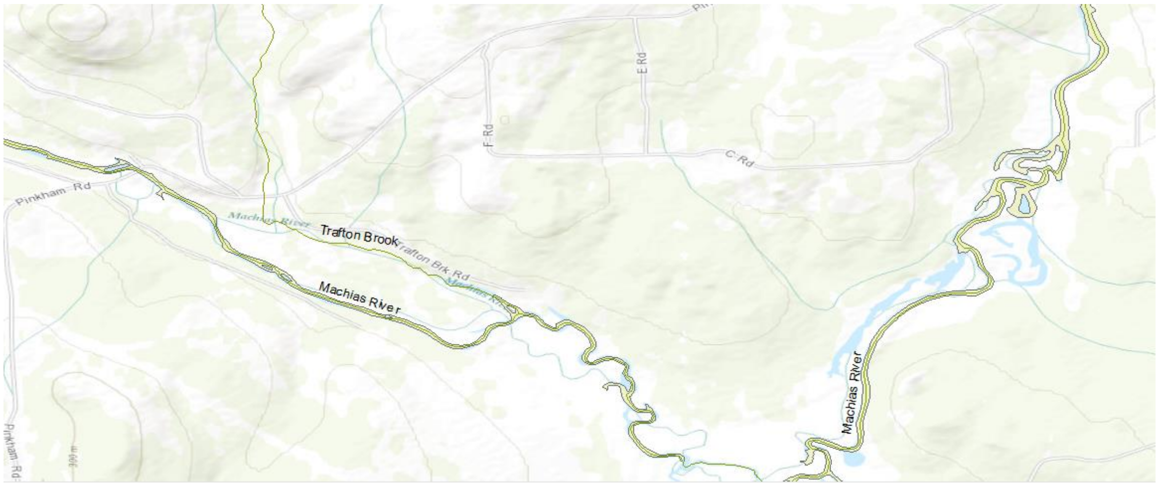


Figure 5.2 Machias River as Line and Polygon Geometry

5.2.1.2 NHDPoint

NHDPoint feature class contains hydrographic and hydrometric features including dams, gaging stations, gates, lock chambers, rapids, rocks, springs, wells, waterfalls and reservoirs. In addition to these land mark features, other locations such as sinkrise and water intake/outflow are also represented. Sinkrise is where a stream disappears underground or where it resurfaces in a karst area and water intake/outflow is a structure through which water enters or exits through a divergence.

Hydrologic features such as springseep, rapids, reservoirs and waterfalls and hydrologic structures such as dams, and gaging stations have multiple spatial footprints as lines and polygons in NHDLine and NHDArea datasets respectively.

5.2.1.3 NHDLine

NHDLine represents some NHDPoint features as linear geometries for cartographic purposes. NHDLine does not participate in the geometric network or assist in identifying flow relationships. Feature types for NHDLine feature class includes bridges, damweir, flume, gate, levee, lock chamber, rapids, tunnel, well and waterfall.

5.2.1.4 NHDArea

NHDArea feature class contains polygon representations of NHDFlowline features such as StreamRiver and Artificial Path. This class represents the areal extent of the water in a stream. Other Area feature types include BayInlet, bridge, canalditch, damweir, flume, levee, rapids, sea/ocean, and submergedstream.

5.2.1.5 NHDWaterBody

NHDWaterbody is a polygon feature class that represents the areal extent of hydrological features that may have been previously represented as NHDFlowline or NHDPoint features along with hydrographic water body features such as lakes, ponds, swamps and marshes.

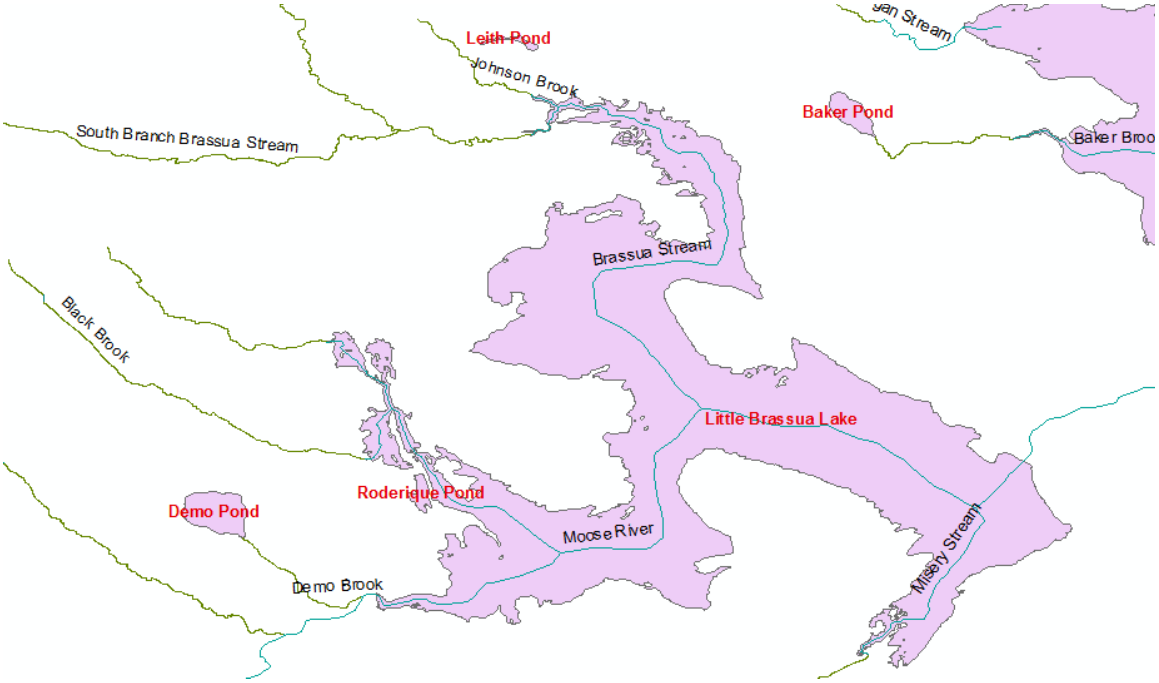


Figure 5.3 Streams Flowing Through Waterbody

5.2.2 Watershed Boundary Dataset

The watershed boundary dataset typically known as ‘Hydrologic Unit’ (http://nhd.usgs.gov/wbd.html), is a polygon feature class which defines the areal extent of the surface water drainage to a point. The Unites States is divided into regions, sub regions, basins, sub basins, watersheds and sub watersheds and arranged hierarchically within each other from smallest unit (sub-watershed) to largest unit (region) (Kapinos et.al 1987). Each hydrologic unit is assigned a Hydrological unit code which is a unique number consisting of 2 to 12 digits depending on the level of classification.

Name	Level	Digit	Number of HUCs	Area Covered (square miles)
Region	1	2	21	177,560
Sub-region	2	4	222	16,800
Basin	3	6	352	10,596
Sub-basin	4	8	2149	700
Watershed	5	10	22000	227 (40,000–250,000 acres)
Sub-watershed	6	12	160000	40 (10,000–40,000 acres)

Table 5.3 Hydrologic Unit Code Classifications

Each hydrologic unit is assigned a hydrologic unit name which is usually the prominent hydrologic feature within the unit. Hydrologic unit boundaries are solely decided based on topography and scientific hydrologic principles. The Watershed Boundary Dataset contains individual feature classes for each level of classification.

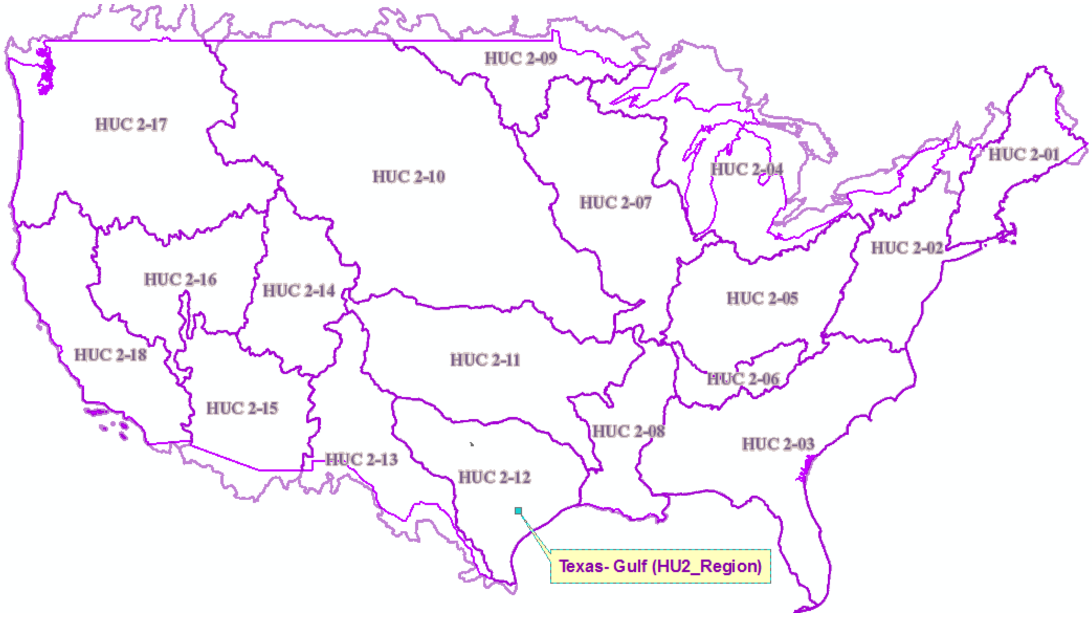


Figure 5.4 Texas-Gulf Region in Hydrologic Unit Map

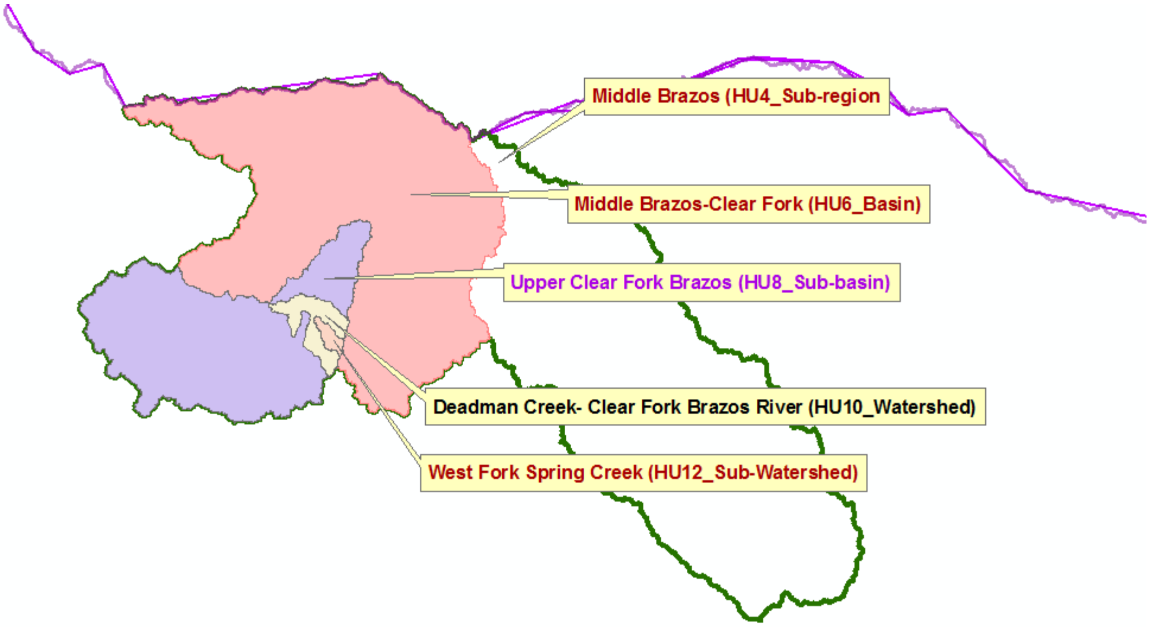

Middle Brazos (HU4_Sub-region
Middle Brazos-Clear Fork (HU6_Basin)
Upper Clear Fork Brazos (HU8_Sub-basin)
Deadman Creek- Clear Fork Brazos River (HU10_Watershed)
West Fork Spring Creek (HU12_Sub-Watershed)

Figure 5.5 Hydrologic Sub Units in Texas–Gulf Region

5.2.3 NHDPlus Dataset

NHDPlus is an integrated suite of application-ready geospatial data products, incorporating many of the best features of the National Hydrography Dataset (NHD), the National Elevation Dataset, and the National Watershed Boundary Dataset (WBD). NHDPlus includes a stream network based on the medium resolution NHD (1:100,000-scale), improved networking, feature naming, and value-added attributes (VAA) (McKay, et al., 2012). The VAAs enhance NHD dataset upstream and downstream navigation capabilities and make it possible to traverse a hydrological network outside of GIS software with SQL queries. FlowlineVAA and PlusFlow tables are the main tables used in this thesis for traversing flowlines and identifying flow relations. The NHD dataset also contains NHDFlow, NHDFlowlineVAA tables that cover the same information as PlusFlow, PlusFlowlineVAA respectively. However NHDPlus dataset contains complete information for flow navigation and hence this thesis uses the NHDPlus dataset predominantly for identifying flow relations. The table schema for the flow tables is similar except for the identifier field. The NHD Dataset uses Permanent_Identifier and NHDPlus dataset used ComID to identify hydrologic features within the dataset.

5.2.4 Flow Relations in NHDFlowline

The NHDFlow table in the NHD dataset and the PlusFlow table in the NHDPlus dataset describe flowing and non-flowing connections between NHDFlowline features. The tables contain entries for: (1) pairs of NHDFlowline features that exchange water, (2) headwater NHDFlowline features, (3) terminal NHDFlowline features, (4) surface water NHDFlowline features that connect to coastline NHDFlowline features, and (5) coastline

NHDFlowline features that connect to each other. Table 5.4 identifies key attributes from these tables that are used to identify flow relationships between features.

Field Name	Description
DeltaLevel	Numerical difference between stream level for From feature and stream level for To feature
Direction	Text or Code to describe direction of flow In – 709 Network start – 712 Network end - 713 Non-flowing -714
From_Permanent_Identifier / FromComID	Identifier of the flowline feature from which the feature flows
To_Permanent_Identifier/ ToComID	Identifier of the flowline feature to which the feature flows

Table 5.4 Flow Table Schema

5.3 Hydro-Semantics of Topological Relations

This section describes how the topological relations between point, line, directed line, and region summarized in section 2.2 apply to hydrologic features represented as point, line and polygon geometries and attaches a semantic meaning to the topological relations in the context of hydrology.

5.3.1 Stream-Stream Relationships

The NHD and NHDPLUS represent streams as directed line segments. In the gazetteer context the unit of interest is a named stream which in the NHD corresponds to an ordered set of connected directed segments with the same GNIS Id and GNIS name. Thus for Stream-Steam relationships, the relationships of interest are between these sets. This section describes how these set relationships map to the topological relations defined by the 9- intersection model for directed line segments. A directed line segment consists of two distinct points, a non-self-intersecting, continuous line that connects the two points, and an orientation imposed on the line, which categorizes the two points as start and end points (Kurata & Egenhofer, 2006).

Among the possible relationships identified by the head-body tail 9-intersection model, few apply given the physical hydrological settings and the constraint of relationships among named streams. The relations based on one intersection that do apply are identified and placed in a hydrological context.

Hydrological networks include complex flowpaths involving streams and waterbodies. Each stream is comprised of multiple reaches with defined flow direction and hence each flowline representing a named stream comprises of multiple line segments with a defined head and tail portion indicating the direction of flow. A constraint of NHD segments is that line segments only intersect at their end points and only in head to tail connections. Assume named streams are the units with names and stream orders. A, B and C represent named stream segments with stream orders A <= B<=C. The connection possibilities are:

A and B have a shared tail location. This split relationship describes two directed lines that coincide at their tails and point in opposite directions. This relationship can be physically realized as two distinct streams emerging from one source (Figure 5.6.)

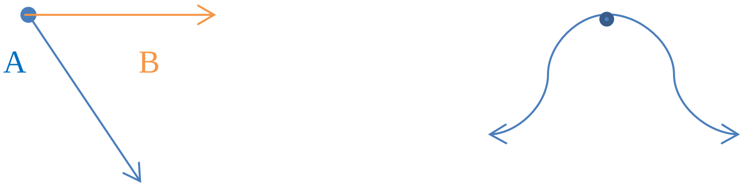


Figure 5.6 Streams from Same Source

B diverges from A (the head of A intersects with tail of B): The diverge/divergedby relations can be physically realized by a distributary of a stream diverging from the main stem of the stream as shown in Figure 5.7

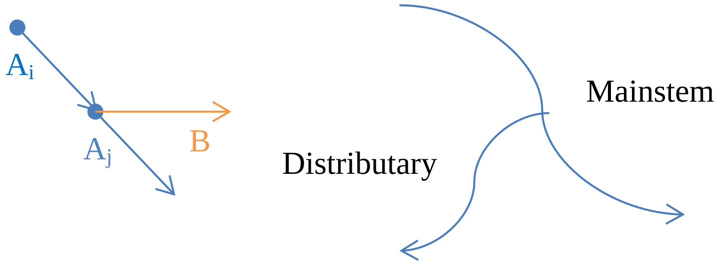


Figure 5.7 Distributary Stream Diverging from Main Stem

A_i precedes A_j (the head of A_i intersects with tail of A_j) and A_j precedes A_k (the head of A_j intersects with the tail of A_k)**:**In this case, there is only one flow path and hence their stream orders will be equal. The precede/follow relations describe the relations between stream segments that form the flow path of the main stem of a stream (Figure 5.8.)

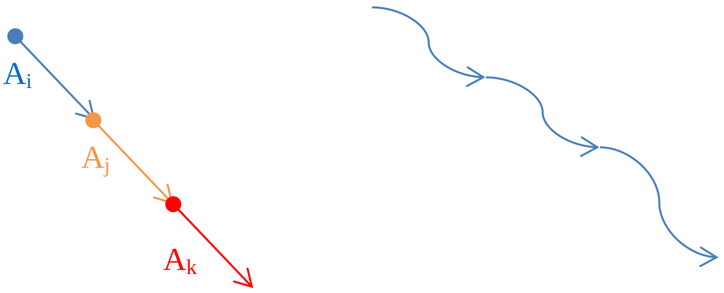


Figure 5.8 Stream Segments Forming a Named Stream Flow Path

In hydrological terms, the preceding stream is upstream of the succeeding stream and hence two cases exist A is UpstreamOf(B) and B is UpstreamOf C. Upstream and downstream relationships also satisfy the transitive property, $P(A,B) \wedge P(B,C), \rightarrow P(A,C)$.

A and B merge and continue as A. The merge/ismergedby relationship describes the tributary relationship if the directed segments carry different names.

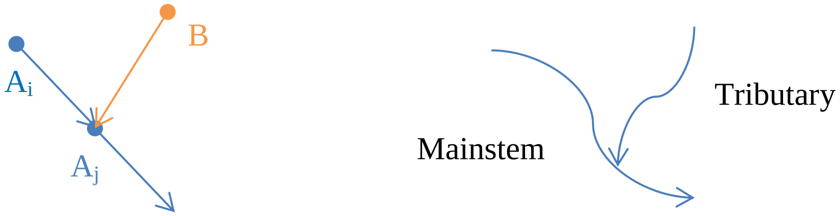


Figure 5.9 Tributary Joining the Mainstem

A and B merge and continue as C. Generally the binary relationship between directed lines "meet" would not apply to two named streams. However two named streams could "meet" and continue as a newly named stream as shown in figure 5.10.

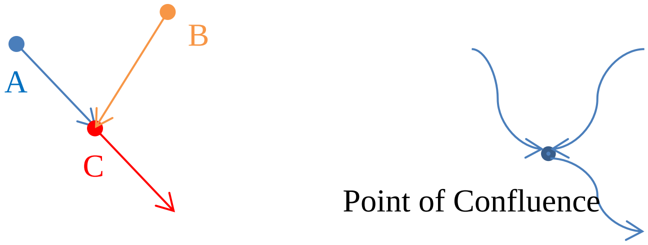


Figure 5.10 Two Streams Meet at a Confluence

Two possibilities of local divergences when one stream diverges from the main path but rejoins the mainstem further downstream are represented in Figure 5.11

a) B diverges and re-merges with A and continues as A and b) B diverges and re-merges with A and continues as C

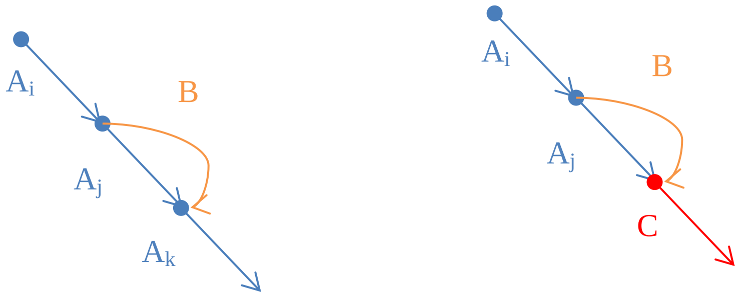


Figure 5.11 Local Divergences that Rejoin the Main Flow Path

5.3.2 Stream - Waterbody

The directed line segment-region relationships (Kurata & Egenhofer, 2007) provide the basis for identifying possible relationships between streams and waterbody features. Cases more specifically that apply to the NHD dataset are:

Case 1: A flowsfrom W. In this case, A exits the region W. The inverse of this relationship is W hasOutflow A.

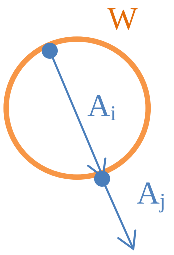


Figure 5.12 Flows From relation between Stream and Waterbody

Case 2: A flowsInto W that is A enters the region W. The converse of this relationship is W hasInflow A.

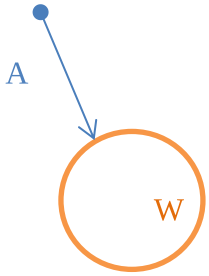


Figure 5.13 Flows Into relation between Stream and Waterbody

Case 3: A flowsthrough W, that is, A enters, crosses and exits region W. If the same named stream enters and exits a waterbody, this relation applies.

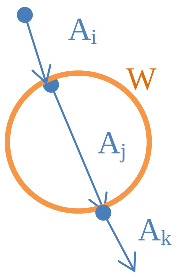


Figure 5.14 Flows Through relation between Stream and Waterbody

Case 4: A flowsthrough W and B flowsInto W. B merges with A within the region W.

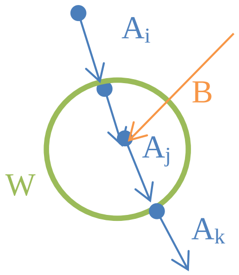


Figure 5.15 Waterbody hasInflow Stream A,B and Stream A flowsthrough W

Case 5: A and B flowsInto W and C flowsFrom W. In this case, A and B merge within the region W.

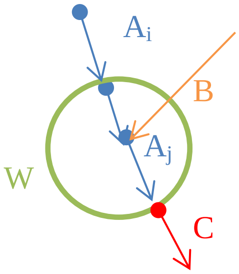


Figure 5.16 Waterbody hasInflow Stream A,B and C as Outflow

5.3.3 Named Waterbody-Waterbody Relationships

While the 9-intersection model and 4-intersection model identify 8 topological relations between two regions (Egenhofer and Franzosa, 1991):disjoint, contains, inside, equal, meet, covers, covered by and overlap, there are fewer possibilities for named waterbody-waterbody relationships. They can be disjoint, in a parthood relation or meet each other. . An example of a meet relation would be ‘Atlantic Ocean meets Pacific Ocean’. Parthood examples include large named lakes which have named bays or coves and ocean with named parts such as Gulfs which in turn may have named parts such as bays, coves, and harbors. For example the Gulf of Mexico has a part named Aransas Bay which itself has a part named Copano BayThese relations meet the core parthood axioms of reflexive $P(x,x)$, antisymetric, $P(x,y) \wedge P(y,x) \rightarrow x=y$, and transitive. $P(x,y) \wedge P(y,z), \rightarrow P(x,z)$. These named waterbodies share some portion of a land water boundary but their waters comingle.

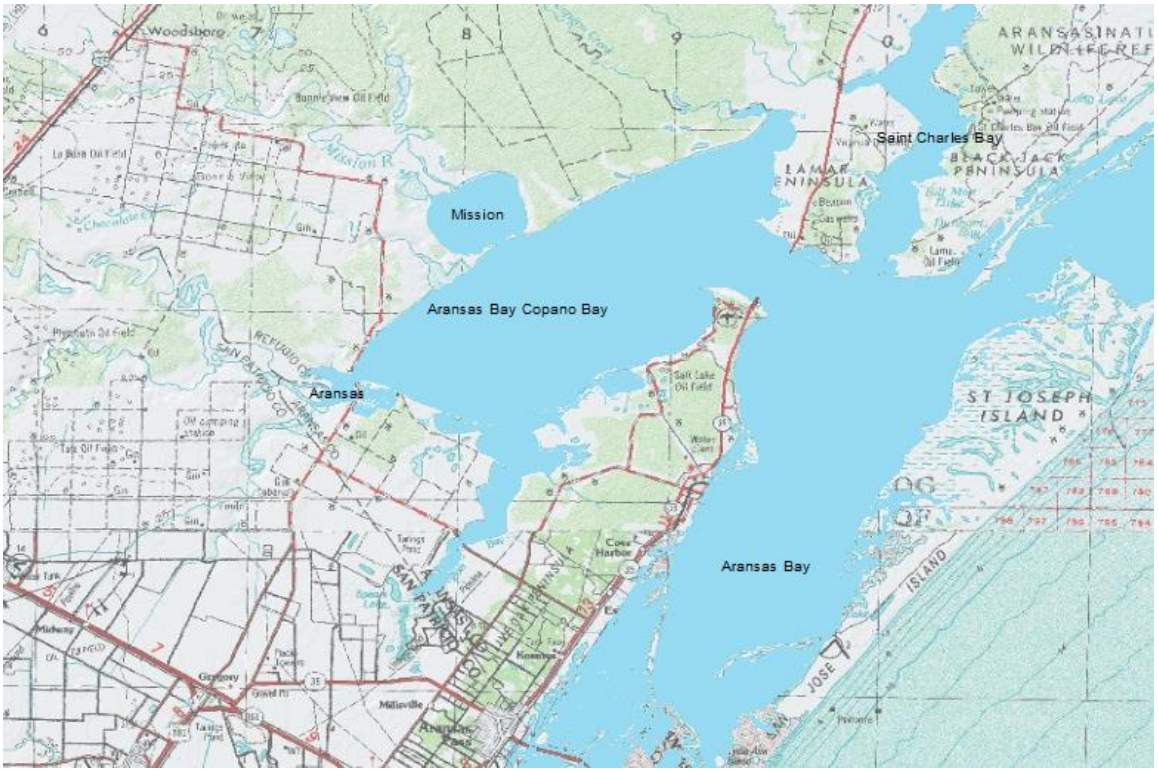


Figure 5.17 Nested Bays in Gulf of Mexico

5.4 Spatial Analysis of NHD Dataset

This section, describes some important attributes in the NHD Plus Value Added Attributes table that are used to make determination of the relationships described above.

The Hydrologic Sequence Number (HYDROSEQ) is a nationally unique sequence number assigned to each flowline segment that places the segment in a hydrologic (flow) sequence. For a given flowline segment, all the upstream segments have a higher Hydroseq number and all downstream segments have a lower Hydroseq number. If flowline segments are processed by hydrologic sequence number in descending order, it is possible to navigate from a stream headwaters and proceed downstream to the stream terminus.

A Level Path Identifier (LEVELPATHI) is a unique identifier for a stream and is assigned to all flowline segments from the stream's mouth to the stream's headwater. The Hydroseq of the flowline at the mouth of the stream is used as the value of the Level Path Identifier.

A Terminal Path Identifier (TERMINALPA) is a unique identifier for all flowlines which flow to the same network terminus (are contained within the same drainage unit). The Hydroseq number of the terminal flowline is used as the Terminal Path Identifier.

A Start Flag (STARTFL) indicates which flowline segments are headwaters and this flag is set to "1" if the flowline is a network start, otherwise it is set to "0".

A Terminal Flag (TERMINALFL) indicates which flowline segments are network ends and terminate at an ocean, Great Lake, Canada or Mexico. This value is set to "1" if the flowline is a terminal flowline otherwise it is set to "0".

5.4.1 Stream Main Stem Identification

The MainStem of the stream can be identified with the LEVELPATHI attribute in the NHDFLowlineVAA table. The LEVELPATHID is the hydrologic sequence number of the terminus flowline in a flow path. All the flowlines with the same LEVELPATHID thus form the mainstem of a stream.

Main Stem Identification Method: Identifies segments that make up the mainstem of a named stream. The method first requires a join of NHDFlowline feature class with PlusFlowLineVAA table based on ComID.

Steps:

1.Select LevelPathID= HYDROSEQ This step identifies unique LevelpathIDs and their frequency of occurrence.

2.For each LevelPathIDs, identify the GNIS_Name and GNIS_ID by selecting features where LEVELPATHID = HYDROSEQ number.

3.For each of the LevelPathIDs and GNIS ID <> “”, select all the flowlines which have the same levelpathids. This selects all the streams which form the main stem of the GNIS named stream.

4.Iterate through the selectionset and generate RDF statements:

a.	gnis_id_stream isMainStempieceOf GNIS_id_mainstemstream

b.	GNISID_id__mainstemstream hasMainStemPiece gnis_id_stream

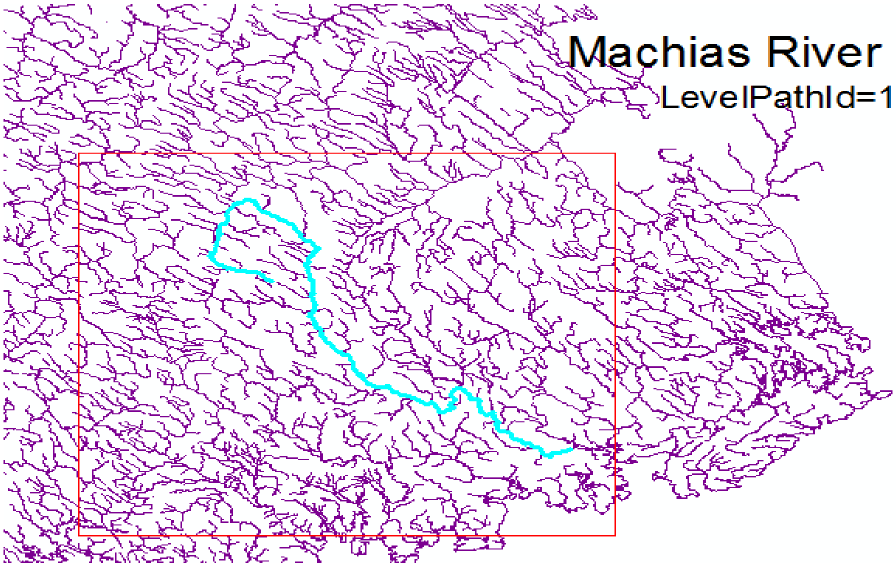


Figure 5.18 Main Stem of Machias River

5.4.2 River Basin

The NHDDataset assigns the same TERMINALPA for all the streams within the same river basin. Hence the name of the river basin and its GNIS_Id can be identified by selecting GNIS_Name and GNIS_Id of the flowline feature, where HYDROSEQ = TERMINALPA. The GNIS_Id of the stream which satisfies the above condition HYDROSEQ = TERMINALPA is assumed as the unique identifier of the basin and is referred as basin_gnis_id in the pseudo code below.

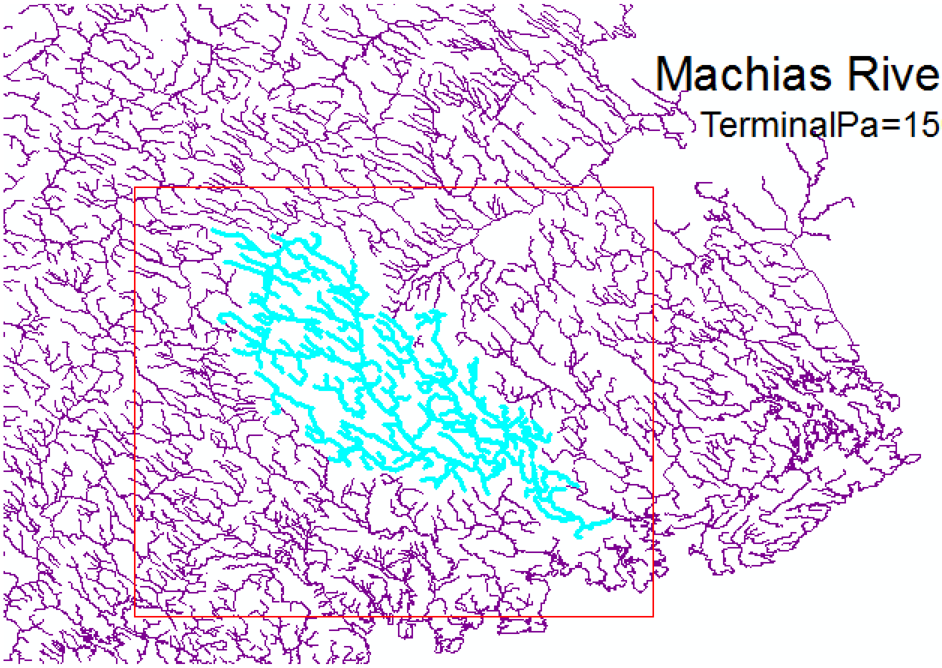


Figure 5.19 Machias River Basin

Input:

NHDFlowline joined with NHDFlowLineVAA table based on ComID.

Steps:

1.Generate Summary statistics for TERMINALPA. This will identify unique TerminalPas and their frequency of occurrence.

2.For each of these TerminalPas and GNIS_ID ,> “”.

3.For each of the TerminalPas, select all the flowlines which have the same TerminalPa.

4.Iterate through the selectionset and generate RDF statements :

a. stream_gnis_id hasRiverBasin basin_gnis_id

b.basin_gnis_id isRiverBasinOf stream_gnis_id

Pseudo-code to extract all the flowlines in a river basin is given below.

```
Generate Summary statistics TerminalPathiSummaryList for the field TERMINALPA
Foreach terminalpa in LevelPathiSummaryList
Get the terminalpa of the desired stream where TERMINALPA = HYDROSEQ
Get all flowlines where TERMINALPA= terminalpa
Foreach flowline :
    Append GNIS_ID to list
Iterate gnis_id list:
    Stream_gnis_id hasRiverBasin basin_gnis_id
    Basin_gnis_id   isRiverBasinOf  stream_gnis_id
```

5.4.3 Tributaries

Tributaries of a particular river can be deduced using Flowline features and the PlusFlow table. By selecting the desired named stream, its tributaries can be identified by retrieving appropriate flow records from the PlusFlow table with terminating flags at the major river.

Input:

NHDFlowline feature class and PlusFlow table

Steps:

1.Select COMIDs of the desired river in NHDFlowline

2.Select all the records in the PlusFlow table where selected ComIDs = PlusFlow.ToComID

3.All the PlusFlow.FromComID's in the selected records from step 2 are the tributaries to the desired river.

4. Iterate through the selectionset and generate RDF statements:

a. GNIS_id.FromComId isTributaryOf GNIS_id.ToComId

b. GNIS_id.ToComId hasTributary GNIS_id.FromComId.

Pseudo-code to extract tributaries for a given NHDFlowline dataset is shown below.

```
Select inRows from NHDFlowline where GNIS_ID is NOT NULL
foreach inRow in inRows:
    Append to list GNISLIST [inRow(comid), inRow(gnis_id)]
for key,value in GNISLIST.items():
    inPlusRows = Get all COMIDs where PlusFlow.ToCOMID =key
    foreach inPlusRow in inPlusRows:
        Retrieve the tributary GNIS_ID from GNISLIST[inPlusRow(comid)]
        Tributary_GNIS_ID isTributaryOf stream_gnis_id
        Stream_gnis_id hasTributary Tributary_GNIS_ID
```

5.4.4 Inflow – Outflow

A lake can have inflow and outflow streams. NHDFlowline has field WBAREACOMI which contains the COMID of an NHDWaterbody, if the flow line passes through the water body. Stream features are connected by Artificial path within the water body polygon. Hence by knowing the flow direction and the association with the water body, it is possible to identify which streams flow into and out of a water body.

Input:

NHDFlowline feature class, NHDWaterbody feature class and PlusFlowTable.

Method:

1. In order to determine the inflows and outflows of a desired waterbody, perform a many-to-one spatial join between NHDWaterbody and NHDFlowline dataset with the spatial condition being ‘boundary touches’. This operation joins all the flowlines whose boundary touches or intersects the boundary of a waterbody.

2. The next step is to determine the flow direction of the streams from the PlusFlow table. For each of the flow lines, determine the preceding (FromCOMID) and succeeding (ToCOMID) flowline identifier.

3.If FromCOMID and ToCOMID are inside the waterbody or FromCOMID and ToCOMID are outside the waterbody, then it can be determined that the stream_A flowsthrough the Lake_B

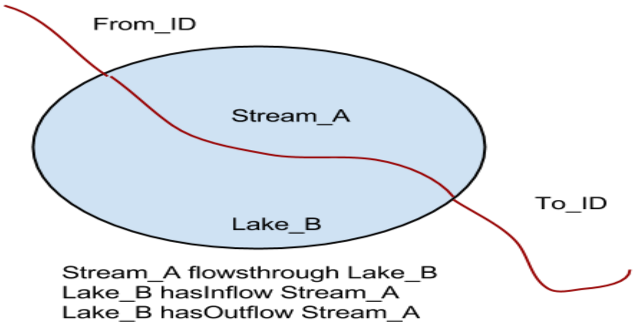

From_ID
Stream_A
Lake_B
To_ID
Stream_A flowsthrough Lake_B
Lake_B hasInflow Stream_A
Lake_B hasOutflow Stream_A

Figure 5.20 Flows Through, HasInflow,and HasOutfow

If FromCOMID is outside the waterbody and ToCOMID is inside the waterbody, it can be determined *that Stream_A flowsInto Lake_B* and its inverse is *Lake_B hasInflow Stream_A.*

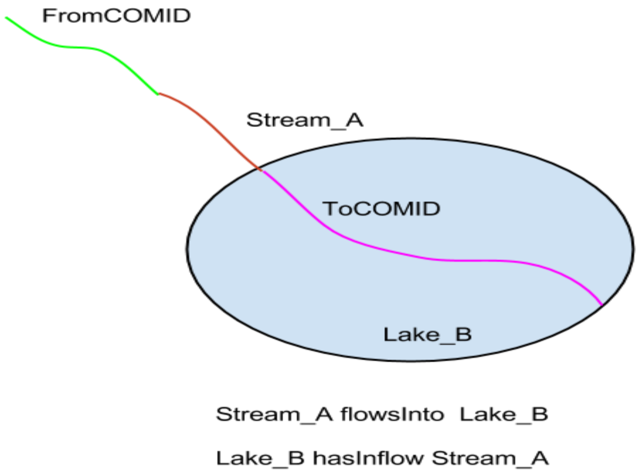


Figure 5.21 HasInflow and FlowsInto

If FromCOMID is inside the waterbody and ToCOMID is outside the waterbody, it can be determined that *Stream_A flowsFrom Lake_B* and its inverse is *Lake_B hasOutflow Stream_A*.

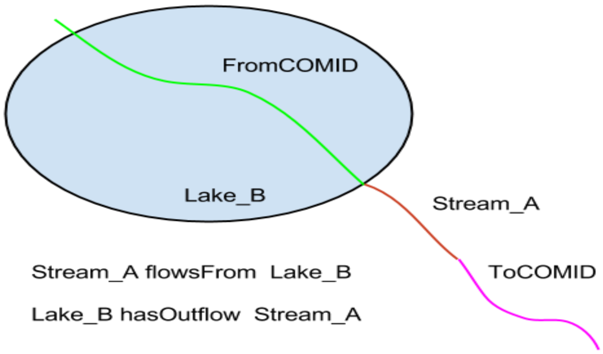


Figure 5.22 HasOutflow and FlowsFrom

```
Get all Flowline_COMIDs from NHDFlowline where WBAREACOMI =  lake_comid
SpatialJoin NHDWaterbody and NHDFlowline where join_condition = "Boundary Touches"
ForEach row in Spatial_JoinOutput:
    Get the flowline_COMID
    Preceding ID=Get preceding flowline id  PlusFlow.FromCOMID where PlusFlow.ToCOMID = flowline_COMID
    Succeeding_id=Get succeeding flowline id  PlusFlow.ToCOMID where PlusFlow.FromCOMID = flowline_COMID
    if preceding_id is in Flowline_COMIDs  (Defined in the first step) and succeeding_id not in Flowline_COMIDs :
        Stream_A flowsFrom Lake_B
        Lake_B hasOutflow Stream_A
    if preceding_id is not in Flowline_COMIDs and succeeding_id  in Flowline_COMIDs :
        Stream_A flowsInto Lake_B
        Lake_B hasInflow Stream_A
    else
        Lake_B hasOutflow Stream_A
        Lake_B hasInflow Stream_A
        Stream_A flowsthrough the Lake_B
```

5.4.5 Springs, Dams, Rapids and Falls

Dam features were downloaded from the GNIS Database and imported as a comma delimited file. The dam locations were imported into ArcMap and a feature class representing dam features was generated. A simple intersection between Flowline features containing streams and the dam feature class provided the association between dams and the streams they are located in. Named dam features from NHDLine and NHDPoint feature classes, if any, were also included in the analysis. The resulting intersection feature class was iterated to generate RDF statements *dam_id isDamOf stream_id* and inverse *stream_id hasDam dam_id*.

The same process was repeated for Springs, Falls and Rapids to determine which flowline features have Falls, Springs and Rapids associated with them.

The flow chart below describes the steps involved in this analysis.

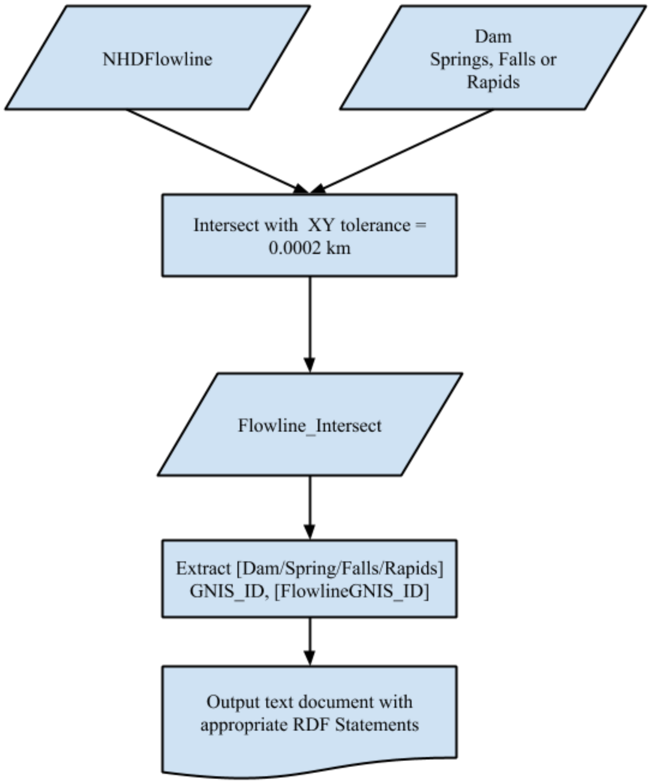

NHDFlowline
Dam
Springs, Falls or
Rapids
Intersect with XY tolerance =
0.0002 km
Flowline_Intersect
Extract [Dam/Spring/Falls/Rapids]
GNIS_ID, [FlowlineGNIS_ID]
Output text document with
appropriate RDF Statements

Figure 5.23 Flowchart for Springs, Dam, Falls and Rapids Features

5.4.6 Nested Watersheds and Nested Bays

Section 5.2.2, described the Hydrologic Unit Codes and how an area is divided and subdivided into hydrologic units. Typically a large watershed *contains* many sub-watersheds and their hierarchy is modelled as a transitive OWL property *hasSubWatershed.*

Input:

Water Boundary Dataset

Method 1:

1. Spatial Join between HUC 8, HUC10, HUC12.

2.Iterate the output of Step 1 and generate RDF statements as HUC8_id hasSubWatershed HUC10_id and HUC10_id hasSubWatershed HUC12_id

Method 2:

1. Relate HUC8 dataset with HUC10 dataset based on the HUC8 unit code.

2.Iterate the selection set for each HUC8 polygon and generate RDF statements as HUC8_id hasSubWatershed HUC10_id.

3. Repeat the steps 1 and 2 with HUC10 and HUC12 datasets.

The NHD Dataset does not have complete spatial footprint information for named Bay features. To overcome this issue, Bay features were downloaded from the GNIS Database and imported into ArcMap as point features. Water boundary dataset has watersheds named after the most prominent feature within the water shed. Hence it is assumed that if a watershed polygon is named as 'Machias Bay', it is considered the polygon representation for 'Machias Bay'. By overlaying the Bay polygon features from the Water Boundary Dataset and the point Bay features from the GNIS database, a subset of nested bay relations can be determined.

Input:

WBD dataset, GNIS_Bay point feature class, NHDArea.

Clip WBD Dataset using NHDArea SeaOcean Feature Type. The resulting clipped feature class contains only coastal features.

Overlay GNIS_Bay point features on the Clipped_HUC feature classes and exclude Bay features in the GNIS_Bay point feature class whose name matches with the WBD HUCName and the point representing the bay is within the polygon representation. This step helps in identifying point features which already have polygon representations, so that an appropriate hierarchy is determined.

Method 1:

1.Spatial Join between Clipped_HUC 8 , Clipped_HUC10,Clipped_ HUC12 and GNIS_Bay features.

2.Iterate the output of Step 1 and generate rdf statements HUC12_id hasSaltWaterBay GNIS_Bay_id. Just a relationship with the finest HUC is needed as the other hasSaltwaterBay relationships can be inferred from the transitive property

Method 2:

1. Relate HUC8 dataset with HUC10 dataset based on the HUC8 unit code.

2. Iterate the selection set for each HUC8 polygon and generate rdf statements as HUC8_id hasSaltWaterBay HUC10_id .

3. Repeat the steps 1 and 2 with HUC12.

4. Overlay GNIS_Bay_feature with ClippedHUC12 and generate rdf statements HUC12_id hasSaltWaterBay GNIS_bay_id

5.4.7 Stream terminal Relations: Source – Mouth

5.4.7.1 Springs as Source

Input:

Generate a point feature class with named feature and featuretype = spring from NHDPoint. Generate a line feature class with Stream named features from NHDFlowLine. Intersect the Spring and stream features with a XY Tolerance of 0.00002 km. The resulting Intersection feature class may contain multiple records for the same Spring feature depending upon the number of stream features that intersect with the spring feature class.

Method:

1. If Spring intersects with only one stream feature and the stream feature is a network start and Spring.SHAPE (Point) is the starting point of

stream.Shape(FirstPoint), then it can be determined that the *Spring isSourceOf stream*.

2. If Spring intersects with only one stream feature (stream_Comid) and Spring.SHAPE is not at the starting point of stream.Shape(FirstPoint), it can be determined that the Spring is not the source of the stream
3. If Spring intersects with two or more stream features and for each stream feature which is a network start and Spring.SHAPE (Point) is the starting point of stream.Shape(FirstPoint), it can be determined that *Spring isSourceOf stream* .

5.4.7.2 Stream as Source

For a given main stem, it is possible to identify the network start using the start flag attribute. Hence to identify headwater flowlines, select the headwater node where start flag = 1. If the headwater node is not a named feature, sort the flowlines based on Hydroseq number and traverse until the most upstream named flowline is reached.

Input:

NHDFlowline joined with FlowlineVAA table based on ComID.

Method:

1. Generate Summary statistics for LevelPathI. This will identify unique LevelpathIs and their frequency of occurrence.
2. For each of these levelpathids, identify the GNIS_Name and GNIS_ID by selecting features where LEVELPATHI = HYDROSEQ number.

3. For each of the levelpathids, select all the flowlines which have the same levelpathids. This selects all the streams which form the main stem of the river.
4. Sort the selection based on HydroSeq number and identify the headwater flowline which has startflag = 1

Pseudo-code to determine if stream is a source of another stream is given below.

```
Generate Summary statistics LevelPathiSummaryList for the field LEVELPATHI
For each levelpath in LevelPathiSummaryList
Get the levelpathi of the desired river where LEVELPATHI = HYDROSEQ
Get all flowlines where LEVELPATHI = levelpathi
Sort flowlines HYDROSEQ descending
Iterate flowline :
    Select name where STARTFL = 1
    If name is not null :
        Generate rdf statement stream_id isSourceOf Mainstem_stream_id
        Generate rdf statement Mainstem_stream_id hasSource stream_id
```

5.4.7.3 Lake as Source

If a water body has a headwater node as its outflow, then it can be determined that the water body is the source of the flowline. If the head water node is not a named feature, then the first downstream feature with a feature name is used.

Input:

NHDFlowline feature class joined with FlowlineVAA table.

NHDWaterbody feature class.

Method:

1. Determine the outflows for a given waterbody.

2. Check if the outflow flowline is a headwater node i.e. STARTFL = 1

3. Generate rdf statements for stream_id hasSource Lake_id and inverse Lake_id isSourceOf stream_id

5.4.7.4 Mouth

The Mouth of a stream can be determined using the FlowlineVAA table and NHDFlowline feature class and Bay feature class similar to the method used to identify river basin features.

Input:

NHDFlowline feature class joined with FlowlineVAA table based on ComID. Coastal feature class with bay features and ocean features.

Method:

1. Generate Summary statistics for TERMINALPA. This will identify unique TERMINALPAs and their frequency of occurrence.

2. For each of these TerminalPas, identify the GNIS_NAME by selecting features where LEVELPATHI = HYDROSEQ number.

3. For each of the TerminalPas, select all the flowlines which have the same TerminalPas and TERMINALFL = 1 (network end).

4. Intersect the selected flowlines with Coastal feature class to identify a river basin and its corresponding mouth which may be a bay or ocean feature.

5. Iterate through the selection set for each river basin and generate RDF statements

   a. Stream. GNIS_id hasMouth coastalfeature.GNIS_id

   b. coastalfeature.GNIS_id isMouthOf Stream.GNIS_id

Pseudo-code to extract stream networks which empty into coastal features.

```
Generate Summary statistics TerminalPathiSummaryList for the field TERMINALPA
For each terminalpa in LevelPathiSummaryList
Get the terminalPA of the desired stream where LEVELPATHI = HYDROSEQ
Get all flowlines where TERMINALPA= terminalPA and TERMINALFL = 1
Intersect with Bay feature class
For each flowline with TERMINALPA = terminalPA
        Generate rdf statements stream_gnisid hasMouth bay_id
        Generate rdf statements bay_id isMouthOf stream_gnisid
```

A similar procedure isrepeated to identify fresh water features such as Lake, Reservoir and Freshwaterbay. The terminal flowline features are intersected with NHDWaterbody feature class which contains lake, reservoir and fresh water bay features.

```
Generate Summary statistics TerminalPathiSummaryList for the field TERMINALPA
Foreach terminalpa in LevelPathiSummaryList
Get the terminalPAof the desired stream where LEVELPATHI = HYDROSEQ
Get all flowlines where TERMINALPA= terminalPA and TERMINALFL = 1
Intersect with NHDWaterbody
```

```
For each flowline with TERMINALPA = terminalPA
        Generate rdf statements stream_gnisid hasMouth waterbody_id
        Generate rdf statements waterbody_id isMouthOf stream_gnisid
```

5.4.8 Upstream and Downstream

Upstream and downstream relations between two flow line features can be determined by building on existing relationships.

a) Streams forming the main stem can be traversed from upstream to downstream.

HydroSeq numbers are used to determine whether a flowline is upstream or downstream to a given flowline feature. Flowline features in the same main stem of a given stream , have their HydroSeq numbers assigned in descending order from the top of the main stem. By sorting HydroSeq numbers in descending order, the flowline features will be sorted from most upstream to downstream order.

```
Generate Summary statistics LevelPathiSummaryList for the field LEVELPATHI
Foreach levelpathi in LevelPathiSummaryList
Get the levelpathi of the desired stream where LEVELPATHI = HYDROSEQ
Get all flowlines where LEVELPATHI = levelpathi
Sort flowlines HydroSeq descending
Foreach flowline :
        Append GNIS_ID to list
Iterate gnis_id list:
         generate RDF statements  gnis_id[i] isUpstreamOf  gnis_id[i+1]
         generate RDF statements  gnis_id[i + 1] isDownstreamTo  gnis_id[i]
```

b) All Tributaries of a stream are upstream to the stream at the point of confluence.

c) PlusFlow table is iterated to determine upstream and downstream flowline features.

CHAPTER 6

IMPLEMENTATION OF HYDROGAZETTEER

This chapter presents a prototype implementation of a hydro gazetteer and various components of the semantically enabled gazetteer along with competency questions that the gazetteer can answer.

6.1 Hydro Gazetteer

The HydroOntology and GazOntology can be viewed as the semantic schema for the HydroGazetteer. These ontologies enable the HydroGazetteer to answer different categories of queries, namely place name queries involving the taxonomy of feature types, queries on relation between named places, and place name queries with reasoning. Figure 6.1 shows an abstraction of queries that can be posed to the HydroGazetter.

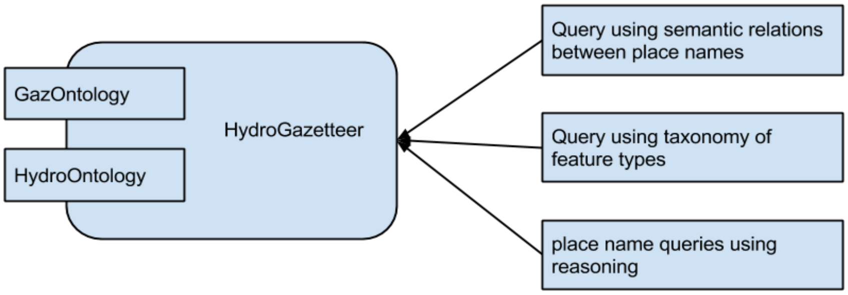

GazOntology
HydroOntology
HydroGazetteer
Query using semantic relations between place names
Query using taxonomy of feature types
place name queries using reasoning

Figure 6.1 HydroGazetteer

The HydroGazetteer was populated with instances of named hydrographic features from the National Hydrography Dataset (NHD) for several watersheds in the state of Maine.

Each feature in the HydroGazetteer is identified by its Geographic Names Information System identifier (GNIS_ID). An OWL dataproperty, gnisName is used to assign the primary name to a feature. The NHD spatial representations for features were extracted to populate the SpatialReferences for each gazetteer entry. Information on tributaries of streams, sources, mouths and other relationships encoded in the HydroOntology was obtained from the NHD database as described in Chapter 5. The point, polyline, and polygon SpatialReference types were encoded as Well Known Text Literals so that the geometries are compatible with GEOSPARQL (Perry & Herring, 2011) queries. Figure 6.2 shows an example of how an instance of a stream is defined in the triple store.

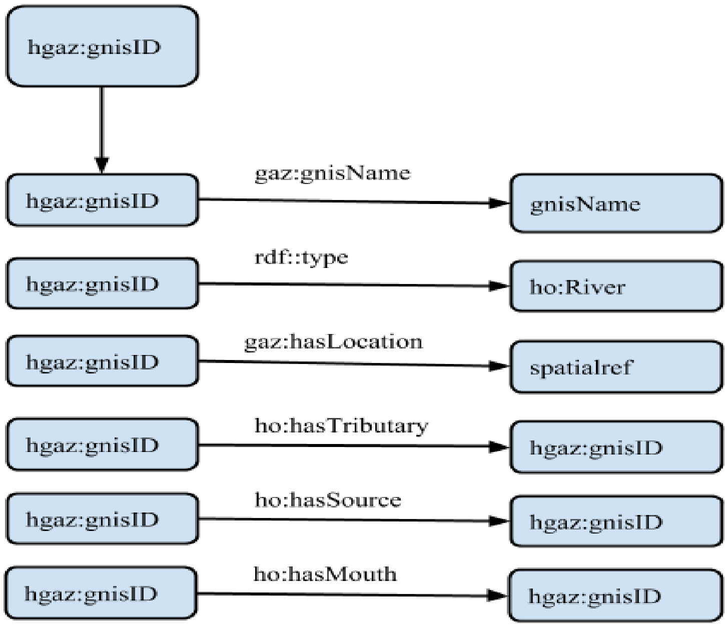


Figure 6.2 Triples Describing a Stream Instance in the Triple Store

6.2 Implementation Components

A high level architecture of the prototype implementation is shown in Figure 6.3. The implementation consists of four components:

- Ontology Design – Tools Used: Top Braid Composer, Protégé.
- Spatial Data Analysis – Tools Used: ESRI ArcGIS, arcpy (ArcGIS Python scripting)
- Triple Store Construction and Querying – Tools Used: Allegrograph
- Mapping Module – Tools Used: Leaflet ESRI Plugin, HTML, JAVASCRIPT

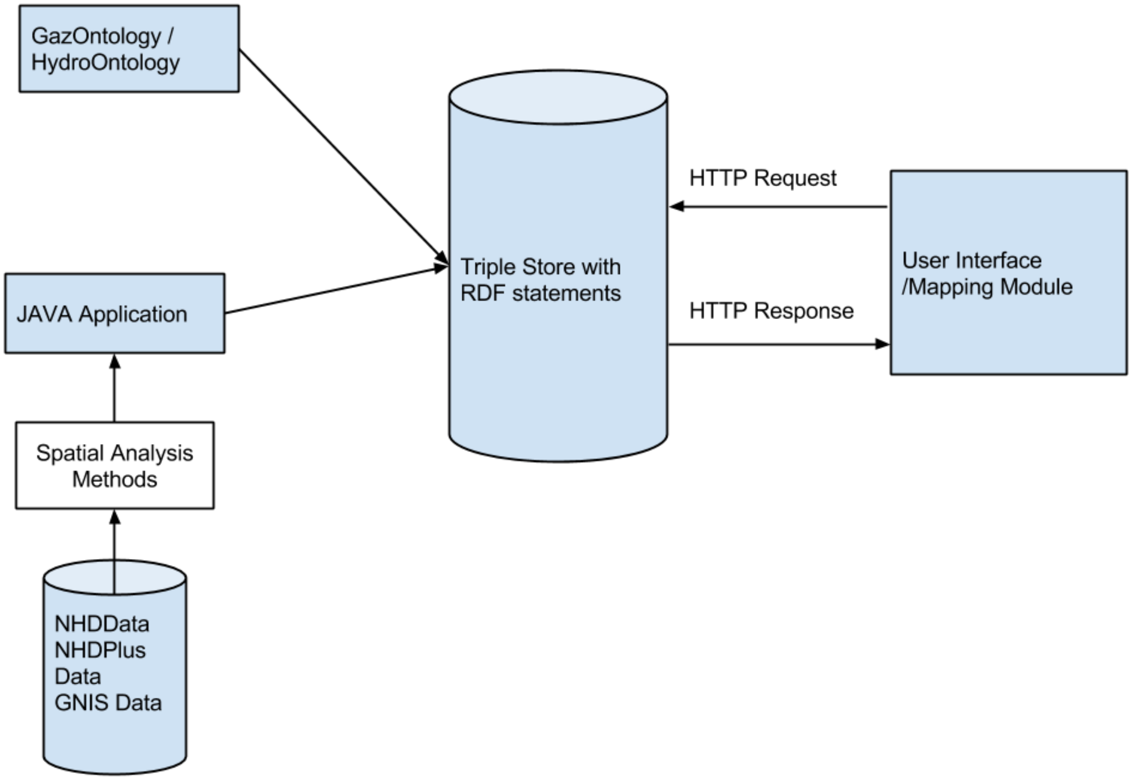
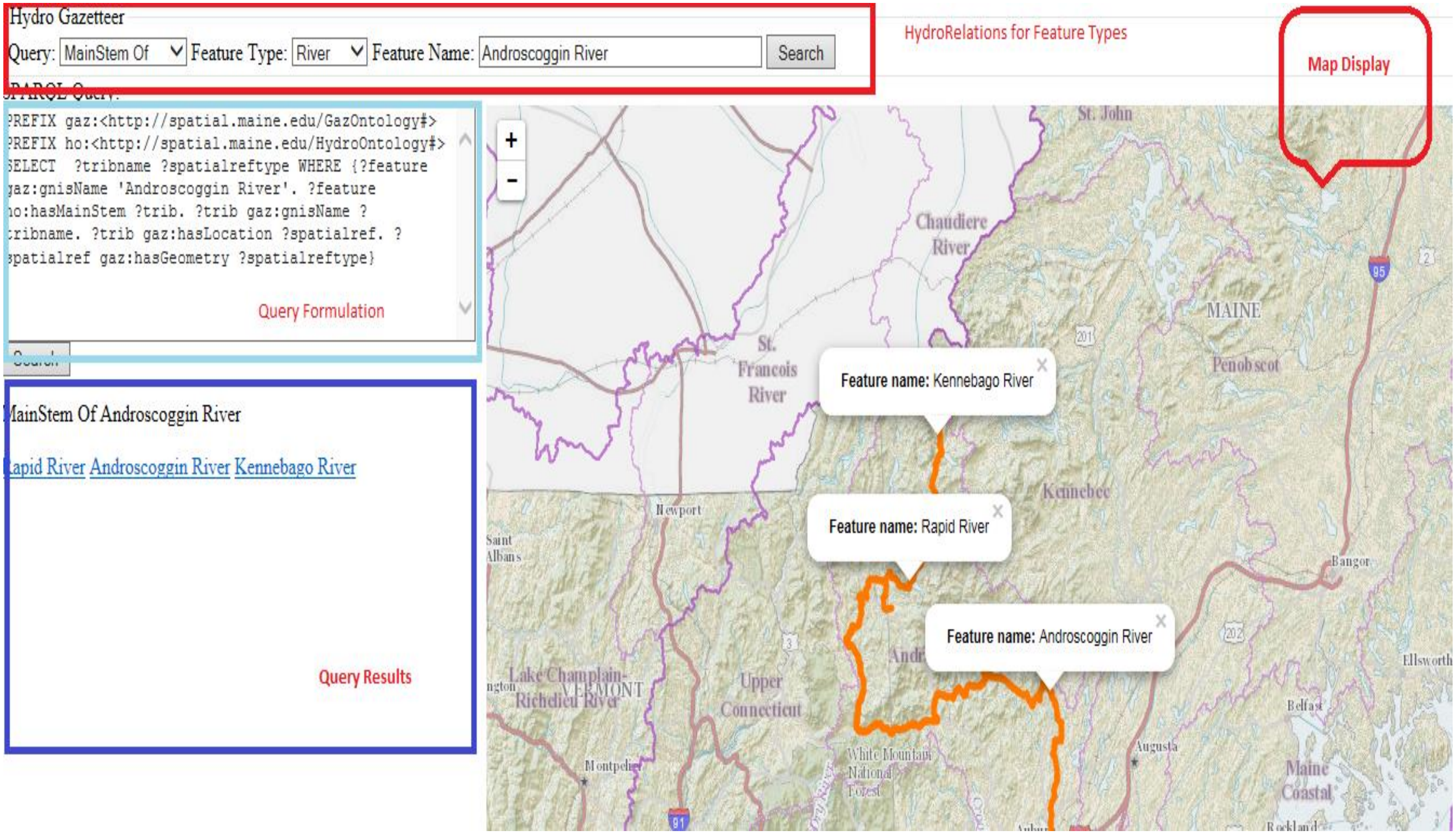


Figure 6.3 General Architecture of the HydroGazetteer Implementation

The ontologies (GazOntology and HydroOntology) were developed using Top Braid Composer and imported into Allegrograph. Spatial analysis methods and python scripts generated text files containing RDF statements that instantiated hydrological features and relationships between them from the NHD datasets. A command line JAVA application using Allegrograph JAVA API was used to parse the RDF statements and populate an Allegrograph triple store. The Allegrograph triple store was selected as it supports spatial queries, inference on RDFS and several OWL properties and SPARQL, the W3C recommended query language for RDF.

The front end of the prototype is a user interface that can be used to query the HydroGazetteer. It was developed as a HTML web page along with Java Script to post SPARQL queries to the Triple Store. Once the triple store was created and populated with hydrologic features and relationships, SPARQL queries can be issued and the results displayed in a web map. SPARQL queries use XML HTTP Request and Response objects to post queries and the results are obtained in JSON format.

The mapping module of the user interface was built with a lightweight mapping component called Leaflet. Leaflet along with the ESRI plugin for Leaflet map control is used to display USGS base map services. The WKT literals that represent the geometry of the hydro feature are transformed to GeoJSON format, which is widely used to display and exchange geographic features in the web. ESRI’s Terraformer javascript package is used to convert a WKT representation to its equivalent GeoJSON format. Figure 6.4 presents the general layout of the Graphical User Interface developed for querying the HydroGazetteer.



Hydro Gazetteer

Query: MainStem Of Feature Type: River Feature Name: Androscoggin River Search

```
REFIX gaz:<http://spatial.maine.edu/GazOntology#>
REFIX ho:<http://spatial.maine.edu/HydroOntology#>
ELECT  ?tribname ?spatialreftype WHERE {?feature
az:gnisName 'Androscoggin River'. ?feature
o:hasMainStem ?trib. ?trib gaz:gnisName ?
ribname. ?trib gaz:hasLocation ?spatialref. ?
patialref gaz:hasGeometry ?spatialreftype}
```

ainStem Of Androscoggin River

apid River Androscoggin River Kennebago River

Figure 6.4 Graphical User Interface Layout

HydroRelations for each of the feature types are listed and presented to the user to facilitate selection. For example, MainstemOf is a predefined relation for selecting the main stem for a queried feature. A user inputs a feature name such as ‘Androscoggin River’ to search for the main stem of this feature. A pre-formulated query is substituted with the user specified place name and this query is posted to the triple store as a HTTP request object. The results are parsed to display the feature name in the text results section. The results of the SPARQL query are returned in JSON format as a part of the HTTP response object. The prototype web application parses the JSON object to retrieve the feature names and their corresponding geometric coordinates. This geometry is then converted to a GeoJSON format and added to the USGS base map to represent the retrieved spatial footprints. This method of overlaying results on the NHD base map service serves the purpose of verifying the results across existing domain data sources in addition to presenting an aesthetic map display.

6.3 Querying the Triple Store - SPARQL queries

(Brank, Grobelnik, & Mladenić., 2005) summarized the current practices and techniques used in ontology evaluation in various fields. Evaluating the use of ontology in an application includes human ability to formulate queries, accuracy of the responses provided by the system’s inference engine and the use of data sources in the domain (such as the NHD Dataset that provides the instances for the classes modelled in the HydroOntology) that the ontology seeks to model. Task-based evaluations provide a framework for assessing the developed ontology and the triple store (Obrst, Ceusters, Mani, Ray, & Smith, 2007). Use cases and scenarios are expressed in the form of competency questions and by answering these questions we demonstrate the capabilities of the developed ontology. The competency questions for evaluating the HydroGazetteer are grouped into four categories based on geometry, hydrologic parts, hydrologic relations and inference queries. Table 6.1 lists the questions from each category in detail and the SPARQL queries and results are presented in the following sections.

Category	Competency Question
Group 1: Retrieve geometry with place names.	1. Retrieve point geometry features – Falls, Springs, Rapids. 2. Retrieve line geometry features – River. 3. Retrieve polygon geometry features – Lake. 4. Retrieve multiple spatial footprints for a given feature X.
Group 2: Retrieve hydrologic parts with place names.	1. Find features that are the sources of a river X. 2. Find feature which is the mouth of a river Y. 3. Find all the stream segments that make up the main stem of river X. 4. Find all the streams in a River Basin R. 5. Find the sub watersheds in a given Watershed W. 6. Find all the direct tributaries of a river R.
Group 3: Retrieve features based on flow relation queries.	1. Find the streams that flow through a waterbody Y. 2. Find all the inflows and outflows of a waterbody Y. 3. Find all the streams that are upstream to a waterbody Y.
Group 4: Inference queries	1. Find all the direct and indirect tributaries to a river R. 2. Find all the bays that are contained within Bay B. 3. Find whether a river R is upstream or downstream to river R1.

Table 6.1 Competency Questions to Evaluate HydroOntology

6.3.1 SPARQL queries based on Place names

Prefixes *hgaz, ho and gaz* are defined for the HydroGazetteer, HydroOntology and GazOntology respectively and are used for the remainder of this section in SPARQL queries.

```
PREFIX gaz:<http://spatial.maine.edu/semgaz/GazOntology#>
PREFIX hgaz:<http://spatial.maine.edu/semgaz/HydroGazetteer#>
PREFIX ho:<http://spatial.maine.edu/semgaz/HydroOntology#>
```

The SPARQL query below retrieves a feature by its name, using the HydroGazetteer property *gnisName*. Any name can be substituted for the object variable and the corresponding GNIS based identifier for the feature will be returned.

```
SELECT ?feature WHERE {?feature gaz:gnisName 'Crystal Spring'}
```

Feature
hgaz:606893

Table 6.2 Crystal Spring Feature ID

The geometry of the feature can be obtained by querying for the spatial reference of the feature and the geometry associated with the Spatial Reference. The *gaz:hasGeometry* property stores the geometry of the feature as a well-known text literal, which can be displayed on a map to indicate the queried feature location. Point representation of ‘Pokey Dam’ is retrieved using the following query and presented on the web map.

```
SELECT ?feature ?spatialref ?geometry
WHERE {?feature  gaz:gnisName ‘Pokey Dam’
          ?feature gaz:hasLocation ?spatialref. (This pattern retrieves the spatial reference.)
          ?spatialref gaz:hasGeometry ?geometry} (This pattern retrieves the WKT geometry.)
```

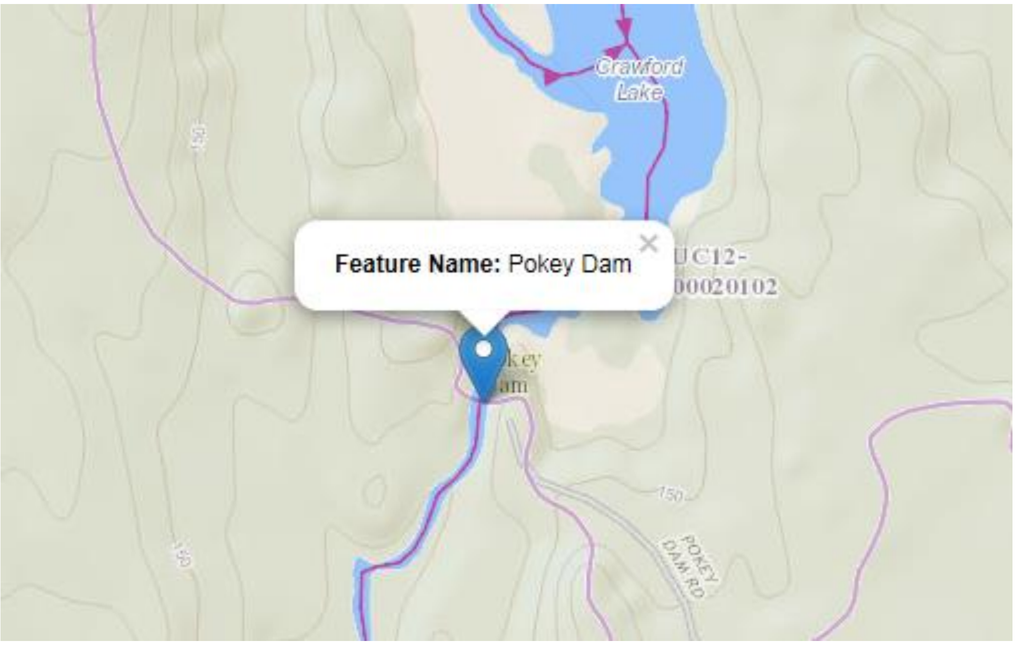
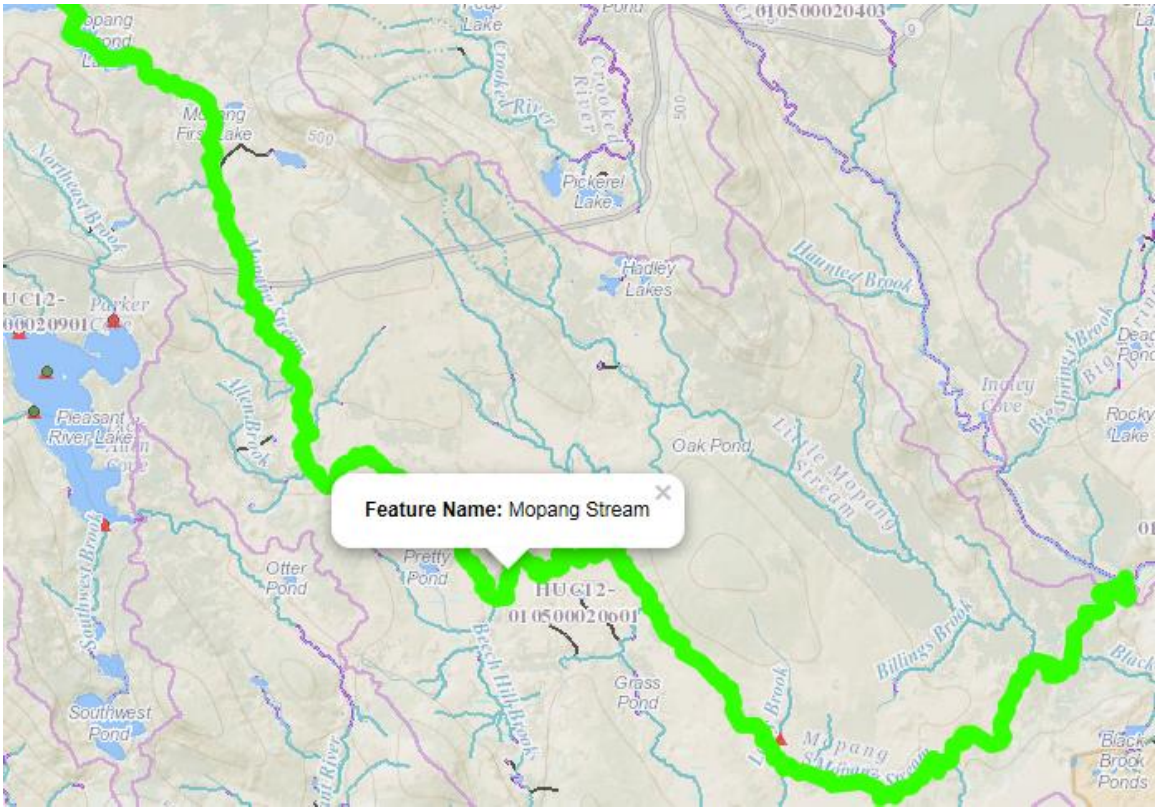


Figure 6.5 Point Representation of Pokey Dam

Streams are represented as lines and the following query returns all the features that match the name ‘Mopang Stream’ along with its coordinates.

```
SELECT ?feature ?spatialref ?geometry
WHERE {?feature  gaz:gnisName ‘Mopang Stream’
          ?feature gaz:hasLocation ?spatialref.
          ?spatialref gaz:hasGeometry ?geometry}
```



Feature Name: Mopang Stream

Figure 6.6 Mopang Stream

Figure 6.7 shows a polygon feature that represents a water body ‘Fifth Machias Lake’.

```
SELECT ?feature ?spatialref ?geometry
WHERE {?feature  gaz:gnisName ‘Fifth Machias Lake’
       ?feature gaz:hasLocation ?spatialref.
       ?spatialref gaz:hasGeometry ?geometry}
```

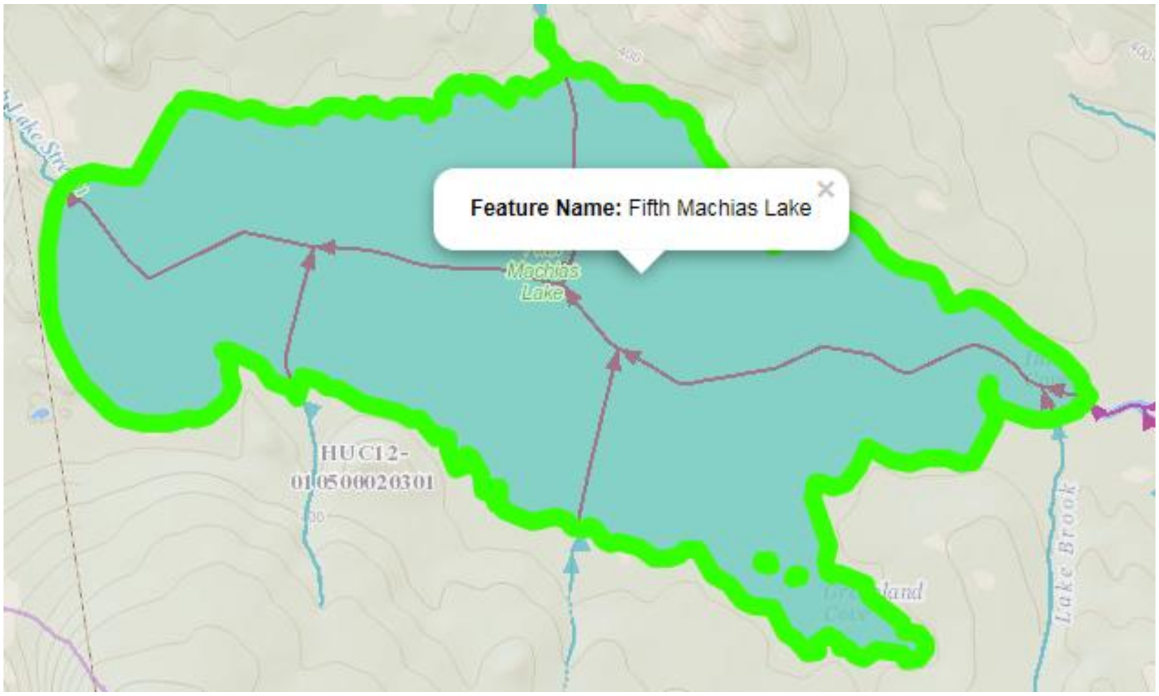


Figure 6.7 Polygon Representation of Fifth Machias Lake

If a feature has multiple footprints in the triple store, all of the spatial references are accessible with the Gazetteer *hasLocation* property. For example, ‘Schoolhouse Rapids’ has two spatial footprints in the HydroGazetteer. The spatial reference for both the representations can be obtained using the following query.

```
SELECT ?feature ?spatialref
WHERE {?feature gaz:gnisName ‘Schoolhouse Rapids’.
            ?feature gaz:hasLocation ?spatialref}
```

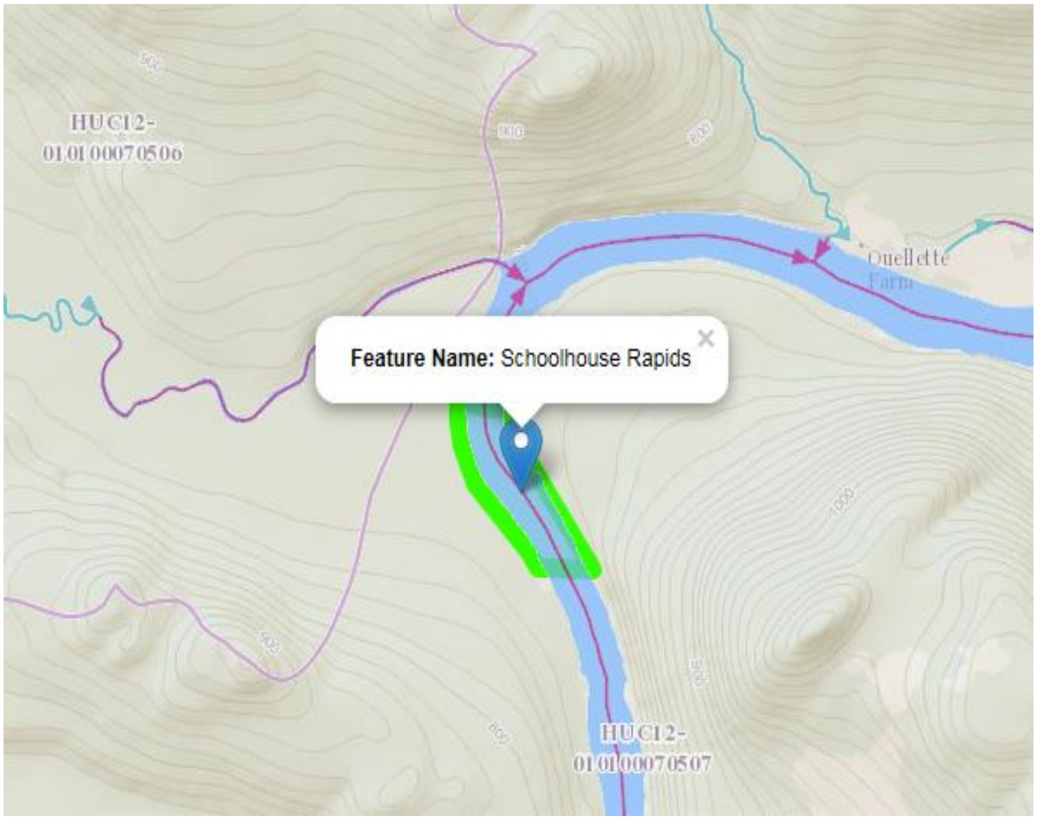


Figure 6.8 Point and Area Representation of Schoolhouse Rapids

If the spatial reference of a feature is known, all the equivalent spatial footprints in different dimensions can be obtained using the *locationEquals* property. For example if the stream is represented as line geometry, locationEquals property can be used to get the polygon representation of the stream feature.

```
SELECT ?feature ?spatialref  ?spatialref1
WHERE {?feature gaz:gnisName ‘Schoolhouse Rapids’.
            ?feature gaz:hasLocation ?spatialref.
            ?spatialref gaz:geomType gaz:Point.
            ?spatialref gaz:locationEquals ?spatialref1}
```

Feature	Spatialref	spatialref1
hgaz:575029	hgaz:point575029	hgaz:polygon575029

Table 6.3 Multiple Spatial References

6.3.2 SPARQL Queries based on Semantics

The main focus of this thesis is to address the place name search problem by explicitly modelling the parts, sub-parts and geographically related properties. The *hasHydrologicalPart, hasHydrologicalRelation* represent the high level relationships among hydrological features as discussed in Chapter 4. This section demonstrates how these relationships can be accessed in SPARQL queries. The SPARQL query below retrieves all the stream names along with the spatial references that make up the main stem of ‘Androscoggin River’.

```
SELECT  ?stemname ?geometry
WHERE {?feature gaz:gnisName ‘Androscoggin River’.
            ?feature ho:hasMainStem ?stem.
            ?stem gaz:gnisName ?stemname.
            ?stem gaz:hasLocation ?spatialref.
            ?spatialref gaz:hasGeometry ?geometry}
```

stemname
Rapid River
Androscoggin River
Kennabago River

Table 6.4 Main Stem of Androscoggin River

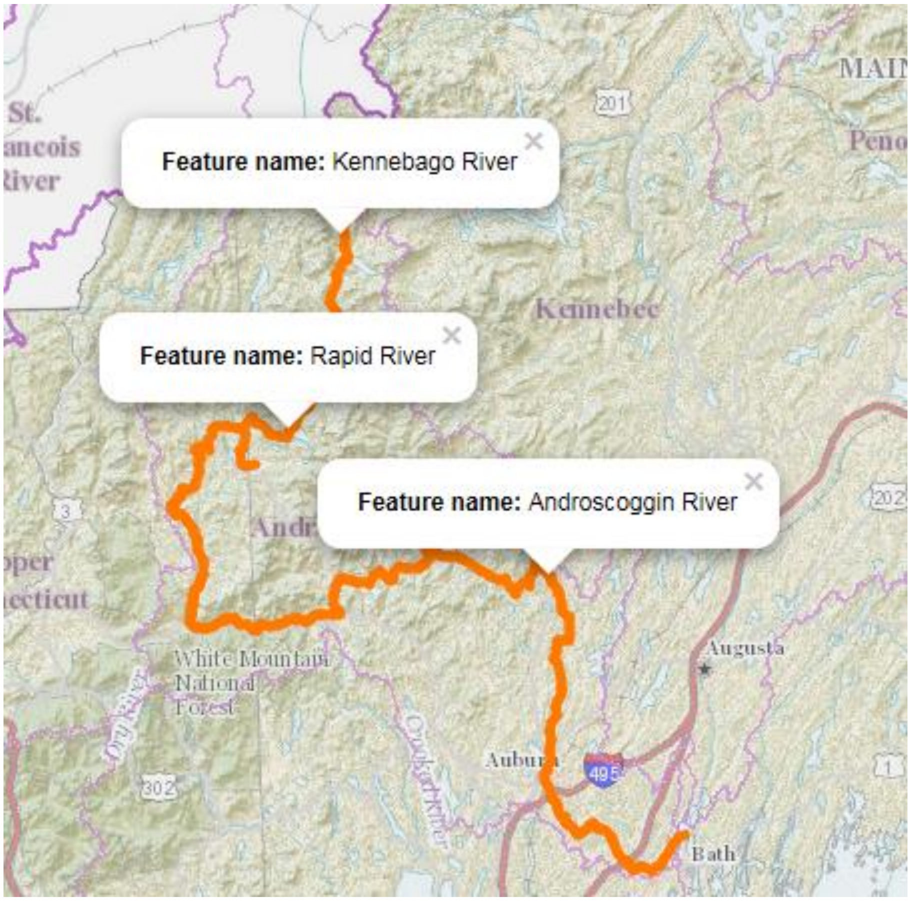


Figure 6.9 Main Stem of Androscoggin River

For example, if the user wants to know the “Main stem of Androscoggin River along with its impoundment structures”, the SPARQL query below retrieves all the streams that are part of the main stem along with dam locations.

```
SELECT ?name ?damname
WHERE {?s gaz:gnisName 'Androscoggin River'.
                ?s ho:hasMainStem ?feature.
                ?feature gaz:gnisName ?name.
                 ?feature ho:hasDam ?damfeature.
                 ?damfeature gaz:gnisName ?damname}
```

Name	Damname
"Androscoggin River"	"Deer Rips Dam"
"Kennebago River"	"Lower Station Dam"
"Androscoggin River"	"Lewiston Falls Project Dam"
"Androscoggin River"	"Pejepscot Dam"
"Androscoggin River"	"Lewiston Falls Dam"
"Androscoggin River"	"Jay Dam"
"Androscoggin River"	"Livermore Falls Dam"
"Kennebago River"	"Big Island Pond Dam Number 4"

Table 6.5 Main Stem of Androscoggin River along with Dams

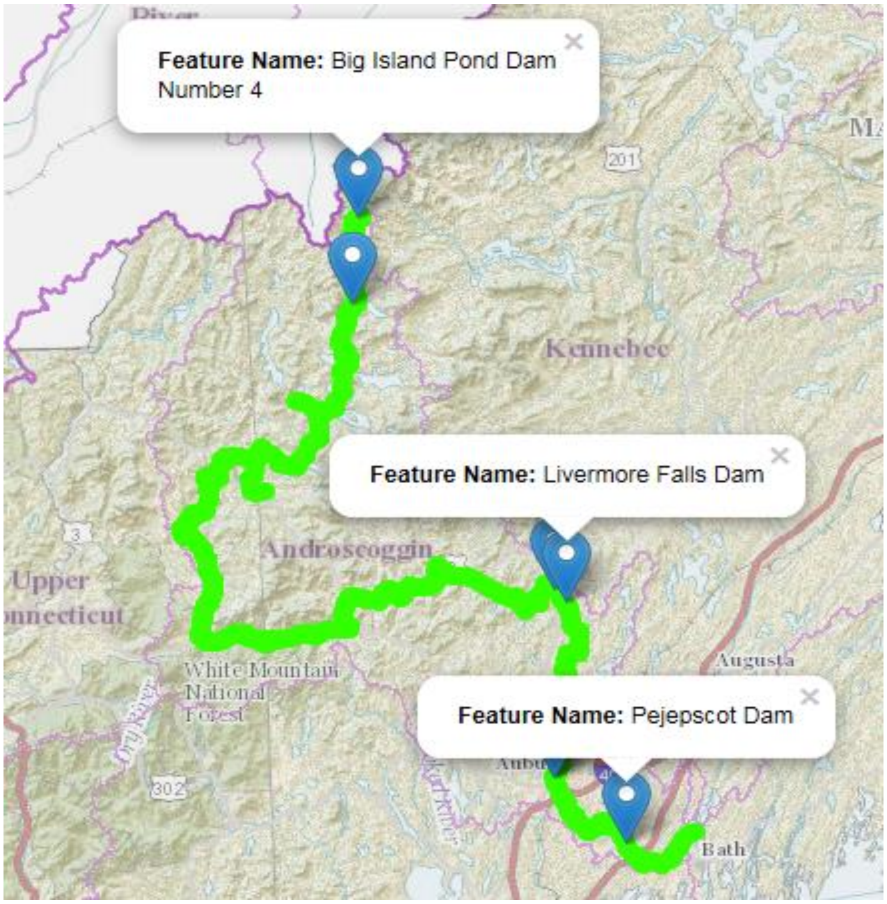

Feature Name: Big Island Pond Dam Number 4
Feature Name: Livermore Falls Dam
Feature Name: Pejepscot Dam
Kennebec
Androscoggin
Upper
nnecticut
White Mountain National Forest
Augusta
Bath
201
302
3

Figure 6.10 Main Stem of Androscoggin River along with Dams

If the user is interested in assessing the effects of a precipitation event in a given area, knowing the tributaries of a major river in the region is necessary. The following query retrieves all the direct tributaries of a given river.

```
SELECT ?tribname ?geometry
WHERE {?feature gaz:gnisName  ‘Mopang Stream’.
   ?feature ho:hasTributary ?trib.
   ?trib gaz:gnisName ?tribname.
   ?trib gaz:hasLocation ?spatialref.
   ?spatialref gaz:hasGeometry ?geometry}
```

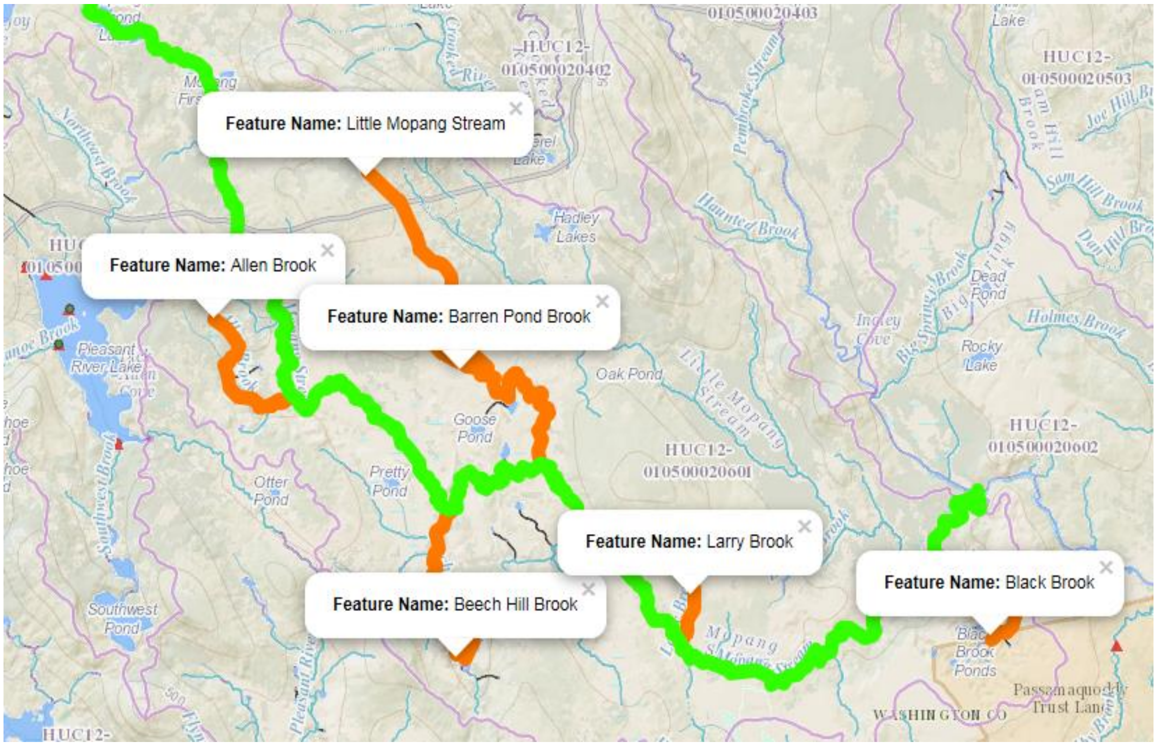


Figure 6.11 Direct Tributaries of Mopang Stream

Since *hasTributary* property is modelled as a transitive OWL property, the entailment rules can be applied to return all the direct and indirect tributaries of “Mopang Stream” by RDFS++ reasoning for a total solution of 8 tributaries.

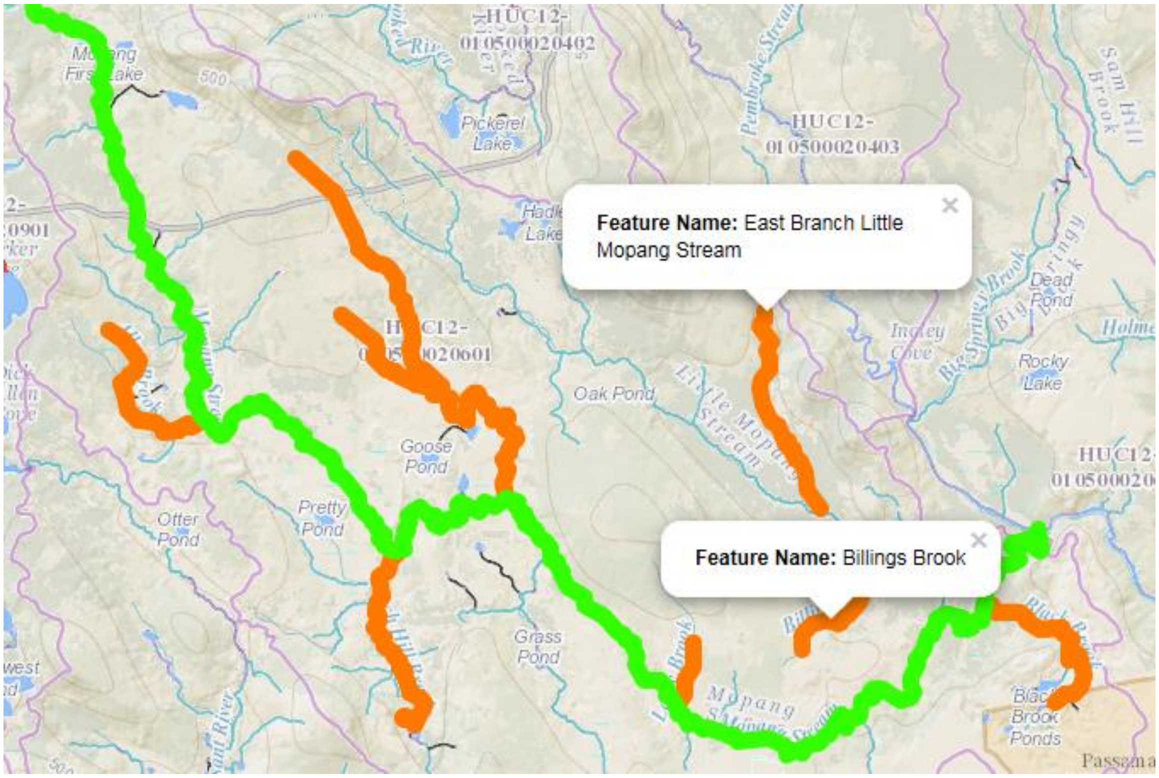


Figure 6.12 Direct and Indirect Tributaries of Mopang Stream

Let us consider the case where the user is interested in ‘Machias River’; however he is not aware of multiple streams with the same name. In this case, the spatial footprint representation of the two streams with the same name help in disambiguating the feature he is interested in. The triple store developed in this work, is capable of returning both streams named “Machias River”, however the actual feature of interest is identified by inspecting the features represented in the map.

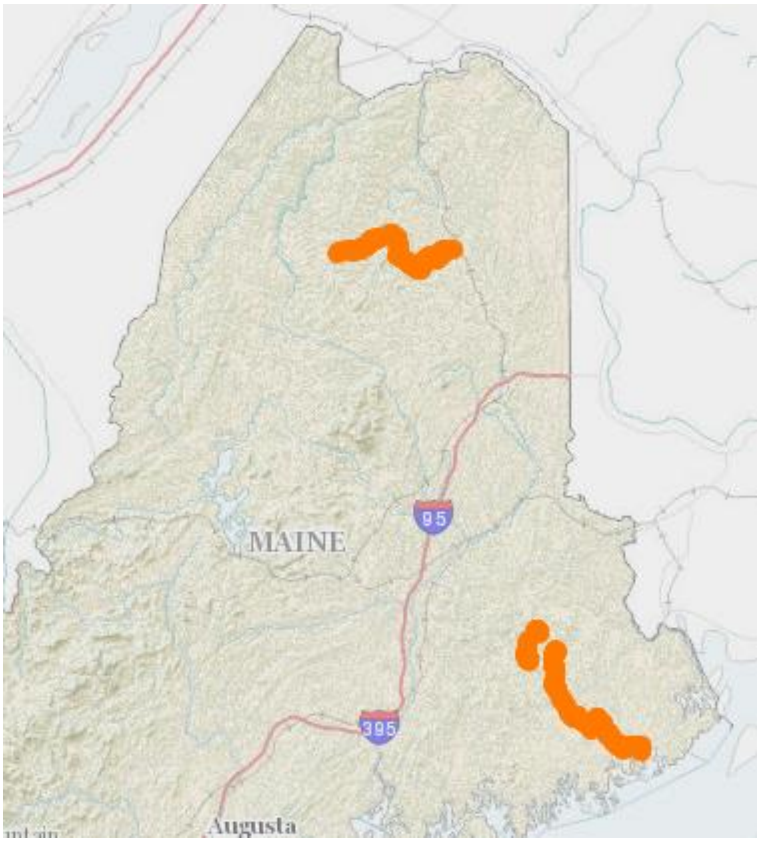


Figure 6.13 Two Streams named Machias River

It is also possible to refine the search for Machias River by expanding the search to include a known hydrological part or a hydrological relation with another feature. Let us say that, the user is interested in the Machias River that flows into the Atlantic Ocean at "Machias Bay" and not "Machias River" which is a tributary of "Aroostook River". The SPARQL query can include an additional graph pattern 'Machias River hasMouth 'Machias Bay' to disambiguate the search for 'Machias River'.

```
SELECT ?feature ?geometry
WHERE {?feature gaz:gnisName 'Machias River'.
              ?feature ho:hasMouth ?s.
              ?s gaz:gnisName 'Machias Bay'.
?feature gaz:hasLocation ?spatialref.
  ?spatialref gaz:hasGeometry ?geometry}
```

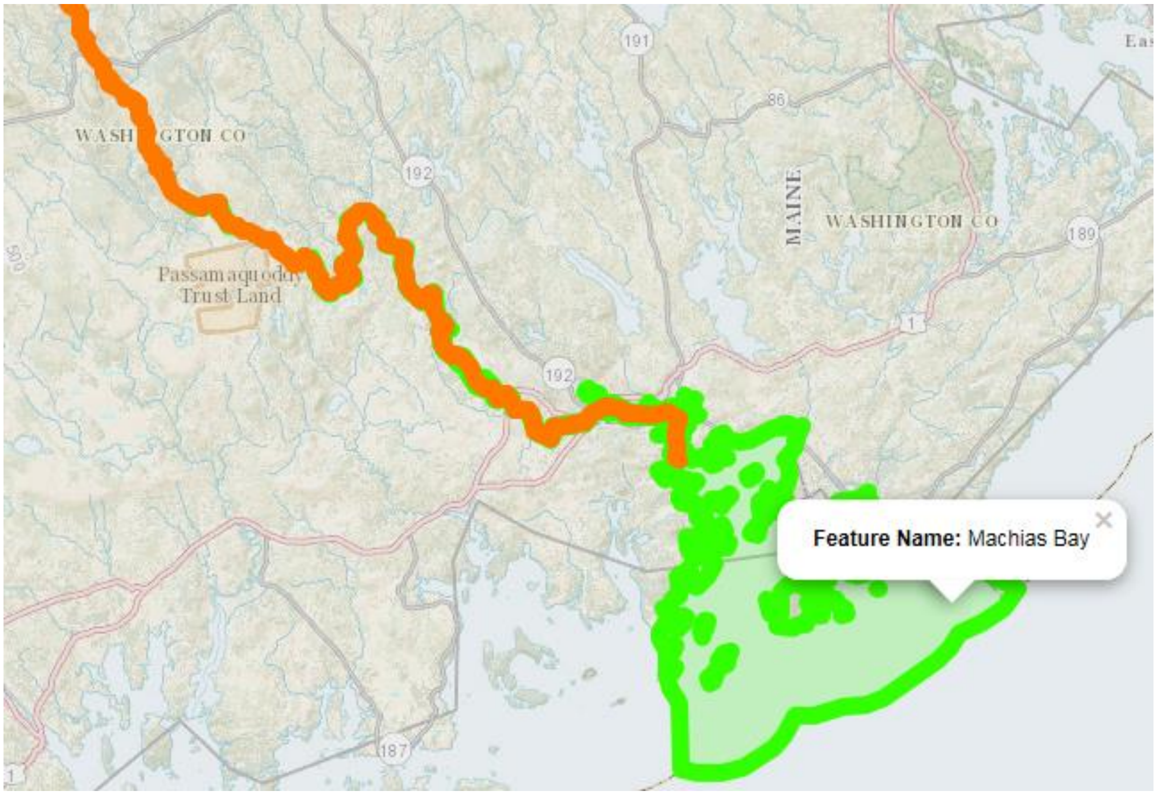


Figure 6.14 Machias River with Mouth at Machias Bay

Inflows and Outflows of a lake are means to identify connected water bodies and the quality of the water body is dependent upon the freshwater inflows and the sediments they bring along. For example, if an invasive species is sighted in a lake or if a toxic product spilled into the lake, by identifying the inflows and outflows, other connected bodies can be identified with the properties *hasInflow/flowsInto,hasOutflow/ flowsFrom* and *flowsThrough*. The SPARQL query below identifies all the streams that flow into ‘Chemquasabamticook Lake’.

```
SELECT ?tribname ?geometry
WHERE {?feature gaz:gnisName ‘Chemquasabamticook Lake’.
    ?feature ho:hasInflow ?trib.
    ?trib gaz:gnisName ?tribname.
    ?trib gaz:hasLocation ?spatialref.
    ?spatialref gaz:hasGeometry ?geometry}
```

Sweeney Brook, Boucher Brook, Fool Brook, Gannett Brook and Ross Inlet are returned as the streams that flow into Chemquasabamticook Lake.

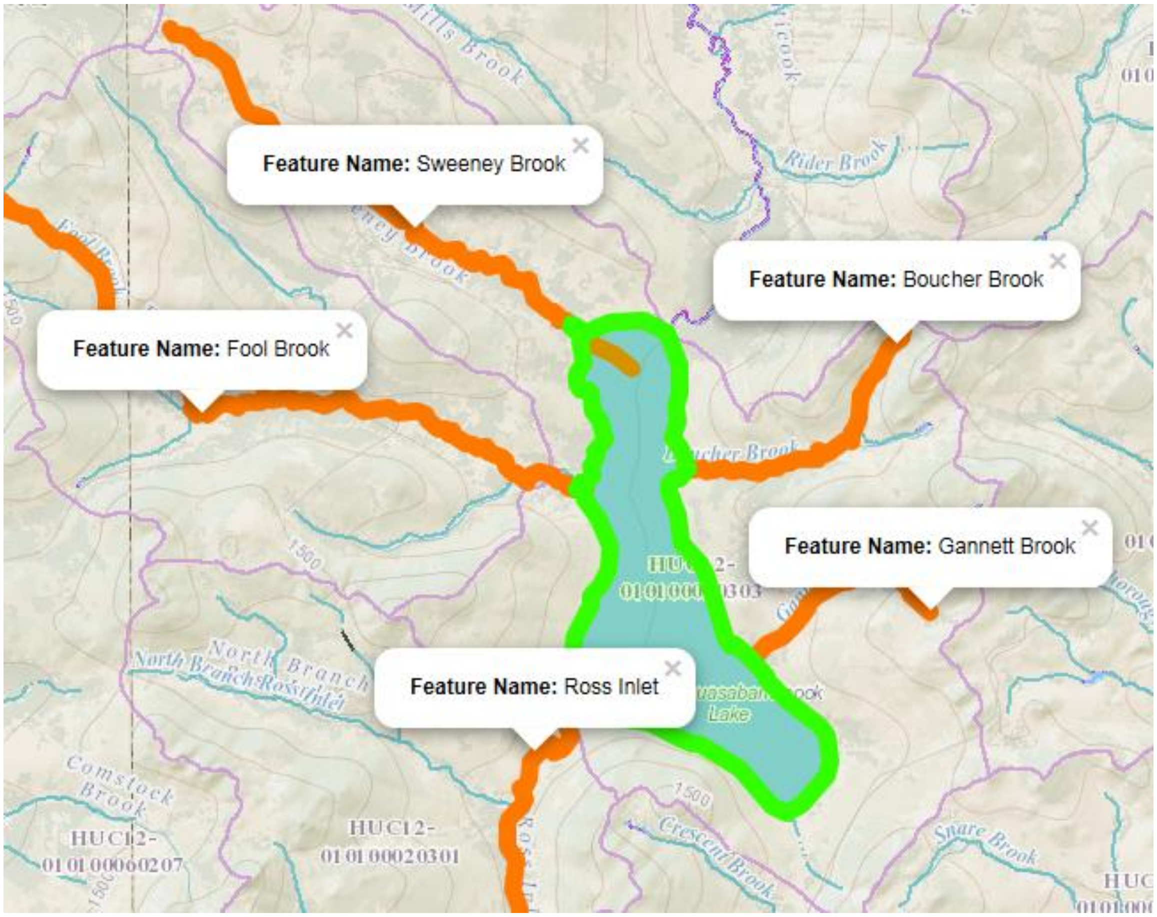


Figure 6.15 Inflows of Chemquasabamticook Lake

Similarly the streams that flow from Chemquasabamticook Lake can be identified by the following SPARQL query.

```
SELECT ?tribname ?geometry
WHERE {?feature gaz:gnisName ‘Chemquasabamticook Lake’.
        ?feature ho:hasOutflow ?trib.
        ?trib gaz:gnisName ?tribname.
        ?trib gaz:hasLocation ?spatialref.
       ?spatialref gaz:hasGeometry ?geometry}
```

Chemquasabamticook stream is shown as the only outflow of Chemquasabamticook Lake with the arrow pointing outside the lake.

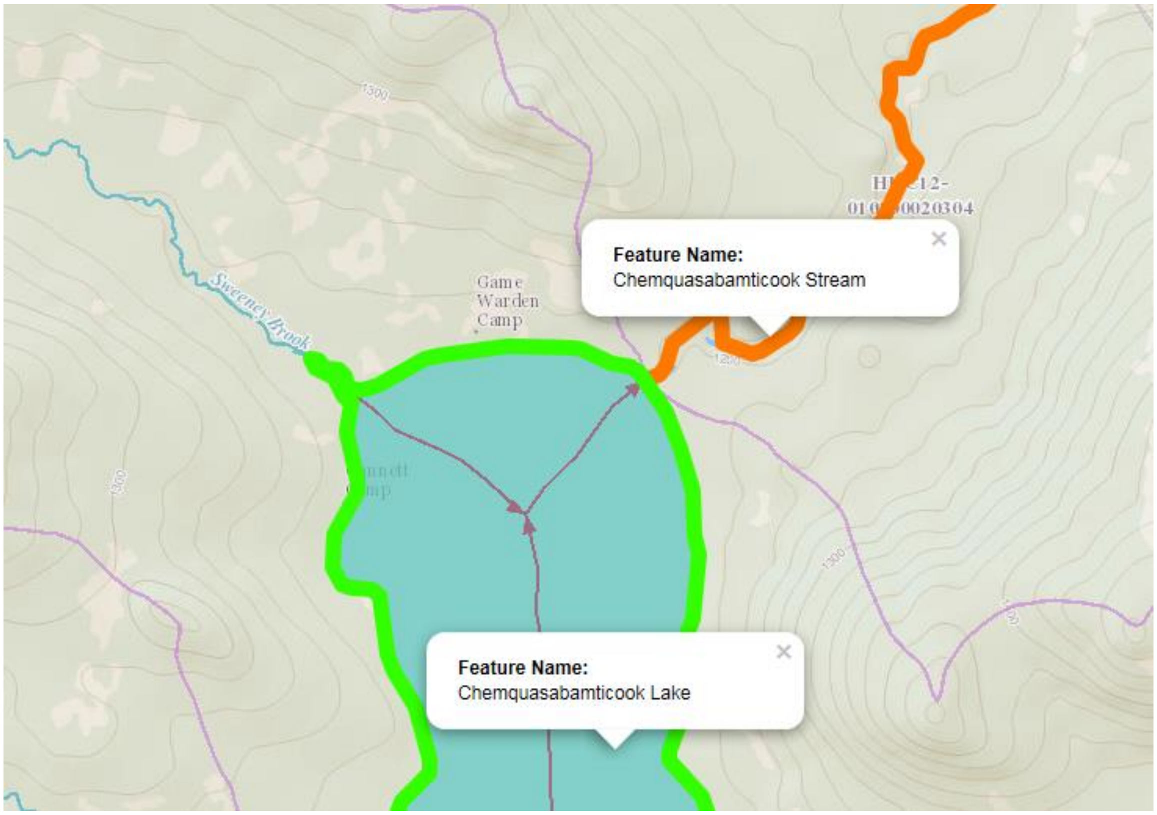


Figure 6.16 OutFlows of Chemquasabamticook Lake

If a user is interested in identifying all the lakes a major river flows through, the flowsThrough relationship retrieves all the Lakes that a river passes through. The following query retrieves all the Lakes that the ‘Allagash River’ flows through.

```
SELECT ?waterbodyname ?geometry
WHERE {?feature gaz:gnisName ‘Allagash River’.
          ?feature ho:flowsthrough ?waterbody.
          ?waterbody gaz:gnisName ?waterbodyname.
          ?waterbody gaz:hasLocation ?spatialref.
           ?spatialref gaz:hasGeometry ?geometry }
```

Waterbodyname
Round Pond
Umsaskis Lake
Long Lake
The Thoroughfare
Harvey Pond

Table 6.6 Allagash River Flows Through Lakes

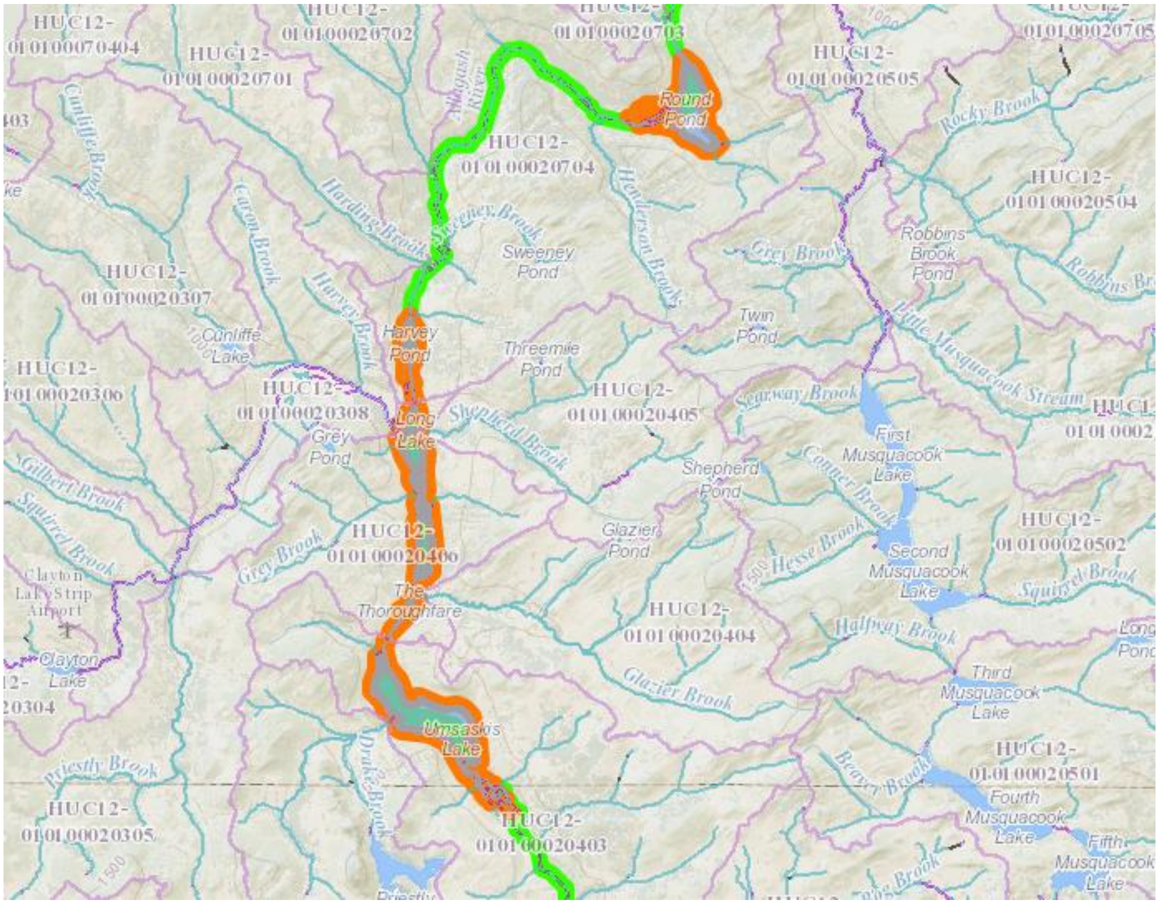


Figure 6.17 Allagash River Flows Through Lakes

Upstream and downstream features can be identified by using the *isUpstreamOf* and *is DownstreamTo* relationships. Both of these relations are modelled as transitive properties and hence RDFS++ reasoning can be applied to determine the upstream or downstream relation between two given streams. The following query identifies the upstream features of the stream ‘Harrow Brook’.

```
SELECT ?upstreamname ?geometry
WHERE {?feature gaz:gnisName ‘Harrow Brook’.
        ?upstreamfeature ho:isUpstreamOf ?feature.
          ?upstreamfeature gaz:gnisName ?upstreamname.
        ?upstreamfeature gaz:hasLocation ?spatialref.
        ?spatialref gaz:hasGeometry ?geometry}
```

Upstreamname
Harrow Lake
Bog Brook

Table 6.7 Upstream Features of Harrow Brook

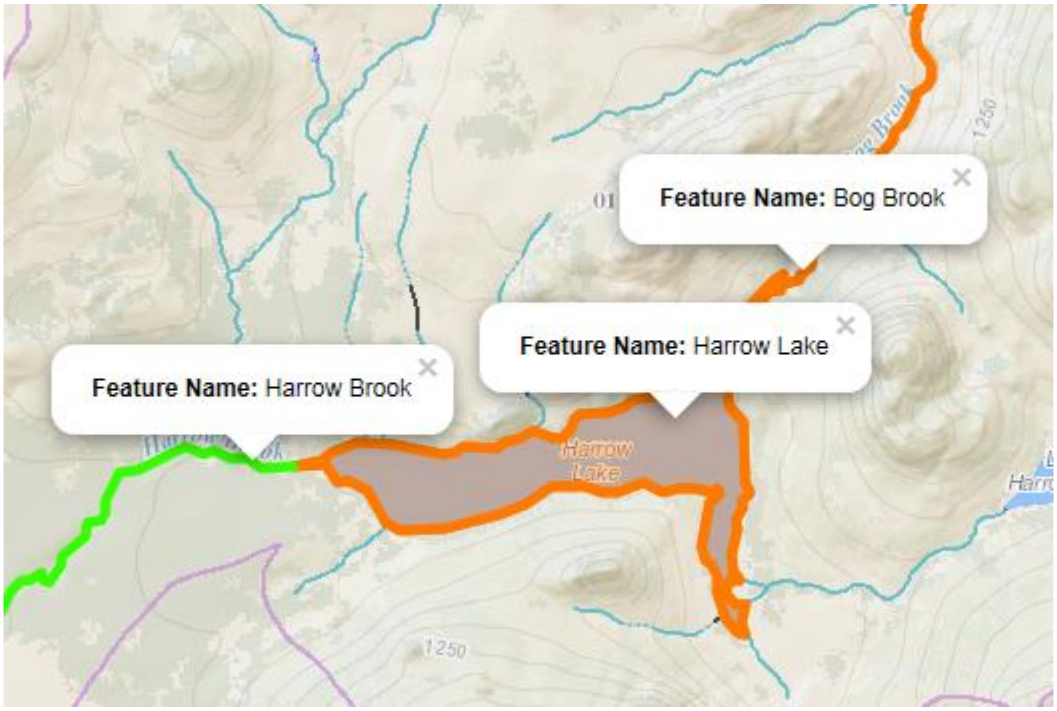


Figure 6.18 Upstream Features of Harrow Brook

The Upstream features of “Mopang Stream” contain about 14 features that are the tributaries, streams and lakes that are upstream of Mopang Stream.

```
SELECT ?tribname ?geometry

WHERE {?feature gaz:gnisName ‘Mopang Stream’.

            ?trib ho:isUpstreamOf ?feature.

              ?trib gaz:gnisName ?tribname.

            ?trib gaz:hasLocation ?spatialref.

            ?spatialref gaz:hasGeometry ?geometry}
```

Tribname
"Allen Brook"
"Mopang Lake"
"Black Brook"
"Mopang First Lake"
"Little Mopang Stream"
"Larry Brook"
"Mopang Stream"
"Beech Hill Brook"
"Mopang Second Lake"
"Barren Pond Brook"
"Black Brook Ponds"
"The Inlet"
"Billings Brook"
"East Branch Little Mopang Stream"

Table 6.8 Upstream Features of Mopang Stream

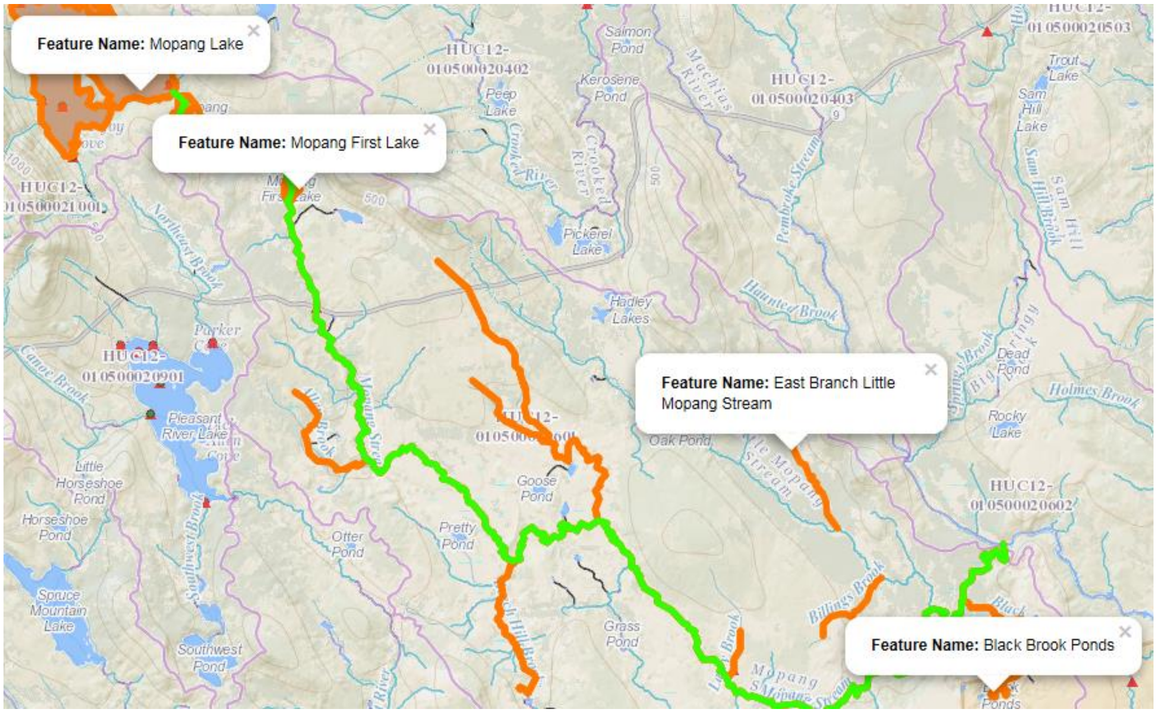


Figure 6.19 Upstream Features of Mopang Stream

The *hasHydrographicPart*, *hasHydrographicRelation* and their sub-properties can be used to deduce the relationship between two given hydrologic features. For example, if a user is interested in identifying how two streams “The Inlet” and “Machias River” are related to each other, we can query the triple store for all the relations if they exist between the queried features.

```
SELECT distinct ?pred
WHERE {?x gaz:gnisName 'The Inlet'.
        ?y gaz:gnisName 'Machias River'.
        ?x ?pred ?y. (This selects the predicates between x and y)
        ?x rdf:type ho:River.
        ?y rdf:type ho:River.}
```

Without any reasoning, The Inlet has a “hasMouth” relationship with “Machias River”. By taking advantage of the inferencing rules, an additional relation “The Inlet isUpstreamOf Machias River” is identified. Along with these sub-properties, the parent properties hasHydrographicPart and hasHydrographicRelation are also retrieved.

It is known that “Allen Brook” is a tributary of “Mopang Stream” and “Mopang Stream” is a tributary of “Machias River” from previous query results. A SPARQL query to retrieve all possible relations between ‘Allen Brook’ and ‘Machias River’ returns the primary relations ho:hasHydrographicPart, ho:hasHydrographicRelation and the sub-properties “Allen Brook ho:hasMouth Machias River”, “Allen Brook ho:isUpstreamOf Machias River” and “Allen Brook ho:isTributaryOf Machias River”.

Sources and Mouth of features can be identified by querying using the hasSource and hasMouth properties. The SPARQL query below identifies the source of “Taylor Branch” stream as “Taylor Brook Pond” which is a Lake.

```
SELECT  ?sourcename ?geometry  ?ftype
WHERE {?feature gaz:gnisName ‘Taylor Branch’.
            ?feature ho:hasSource ?source.
            ?source rdf:type ?ftype.
            ?source gaz:gnisName ?sourcename.
             ?source gaz:hasLocation ?spatialref.
              ?spatialref gaz:hasGeometry ?geometry}
```

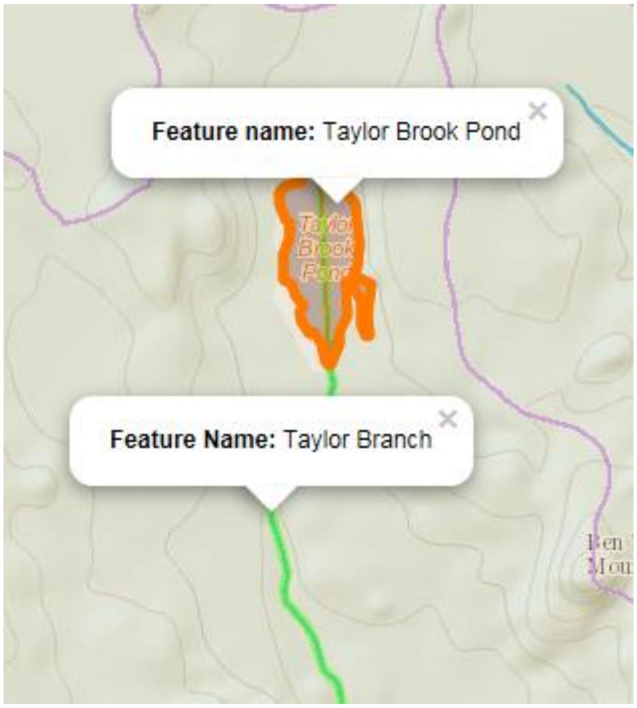


Figure 6.20 Source of Taylor Branch

```
SELECT  ?tribname ?geometry

WHERE {?feature gaz:gnisName ‘Pleasant River’.

   ?feature ho:hasSource ?trib.

    ?trib gaz:gnisName ?tribname.

   ?trib gaz:hasLocation ?spatialref.

   ?spatialref gaz:hasGeometry ?geometry}
```

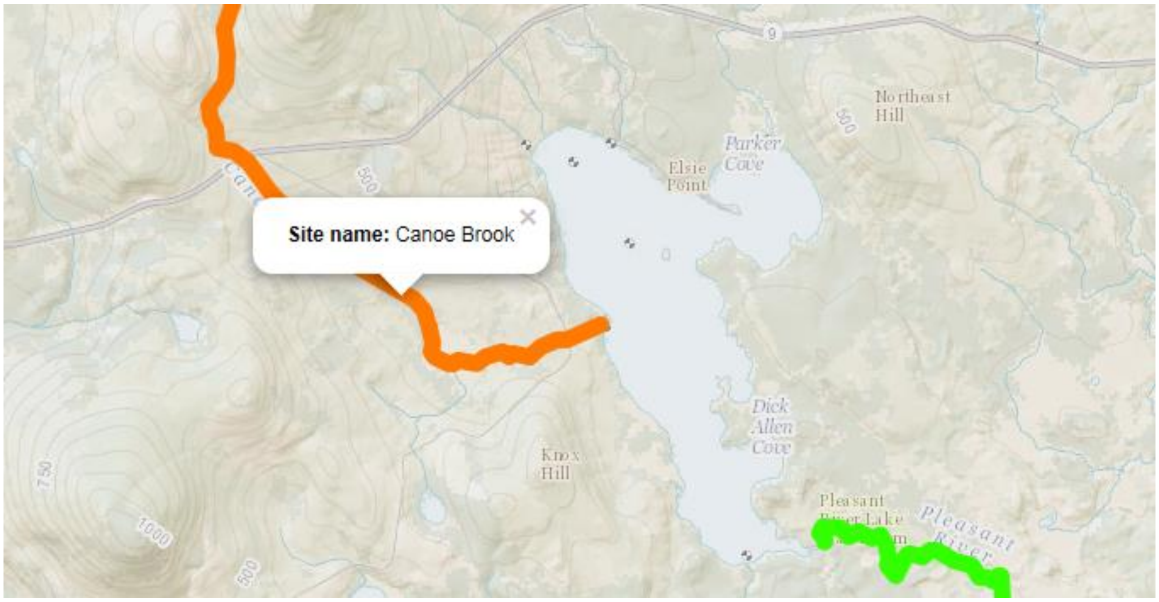


Figure 6.21 Source of Pleasant River

All the streams that terminate at Pleasant River can be retrieved by querying for all the features that are instances of *River* and have a hasMouth relation with “Pleasant River”.

```
select ?tribname ?geometry
where {?s gaz:gnisName ‘Pleasant River’.
          ?feature ho:hasMouth ?s.
          ?feature gaz:gnisName ?tribname.
         ?feature gaz:hasLocation ?spatialref.
           ?spatialref gaz:hasGeometry ?geometry}
```

Tribname
Montegail Stream
Northeast Brook
Knowles Brook
Fred Dorr Brook
Bill Smith Brook
Ingersoll Branch
Marst Brook
Canoe Brook
Bells Brook
Little River
Taylor Branch
West Branch Pleasant River
Beaver Meadow Brook
Branch Brook
Colonel Brook
Northwest Branch Montegail Stream
Western Little River
Pleasant River
Southwest Brook

Table 6.9 Streams that Terminate at Pleasant River

Transitive relations can be used to identify features that completely contain other features, such as watersheds and bays. Bays that contain other bay features such as smaller bays, coves etc. can be retrieved using the hasSaltWaterBay property. The following SPARQL query retrieves all the bays contained within “Frenchman Bay”.

```
SELECT ?name
 WHERE {?x gaz:gnisName 'Frenchman Bay'.
                ?x ho:hasSaltWaterBay ?y.
                ?y gaz:gnisName ?name}
```

Name
Taunton Bay
Preble Cove
Hog Bay
Egypt Bay

Table 6.10 Bays of Frenchman Bay

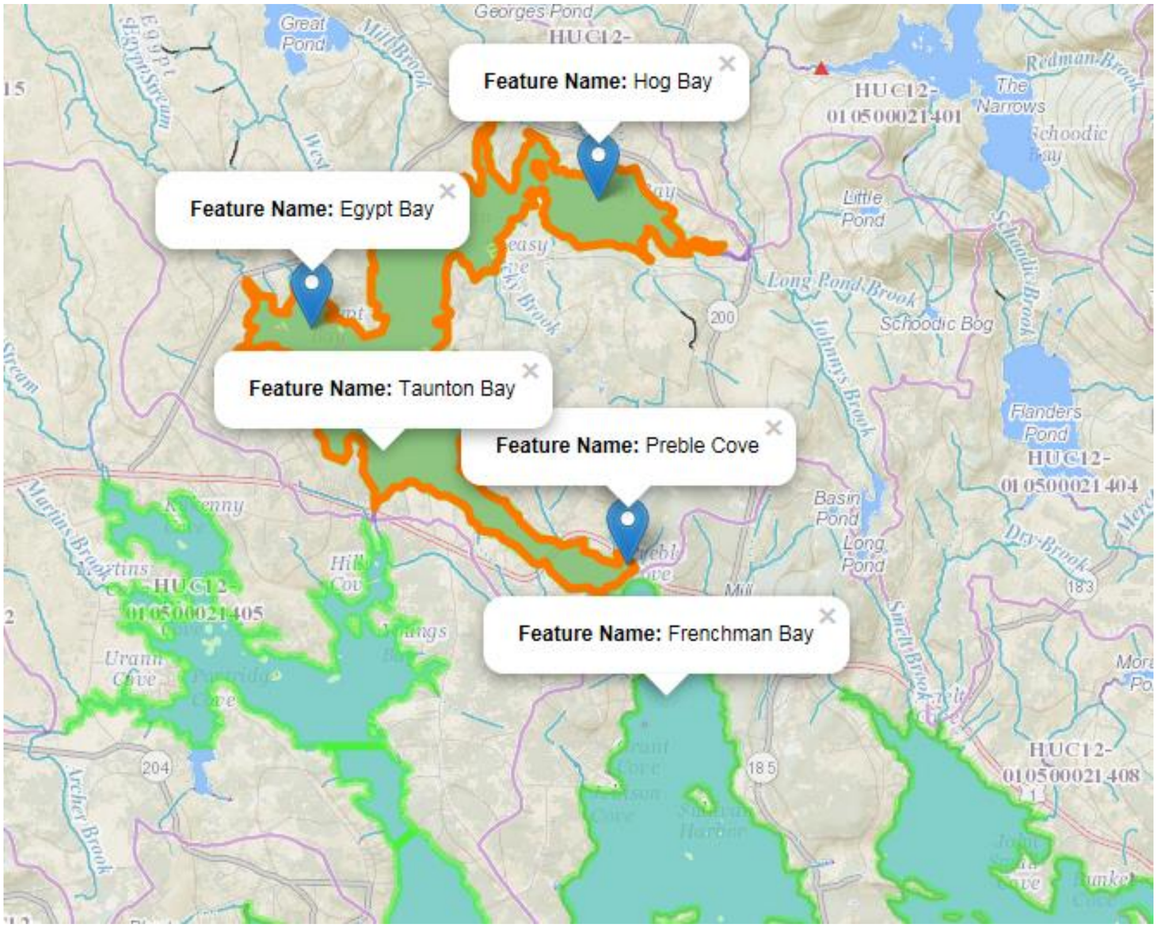


Figure 6.22 Bays of Frenchman Bay

This chapter discussed the overall architecture of the developed HydroGazetteer and the implementation aspects including ontology development, instantiation of the classes modelled and the competency questions that are used to evaluate the developed ontology. SPARQL queries that interact with the triple store were presented along with the graphical representation of the results.

CHAPTER 7

RESULTS AND DISCUSSION

This chapter presents a summary of the thesis along with the results and contributions. The chapter also presents future research that could follow from this work.

7.1 Summary

This thesis presented an approach to enhance place name searches by modelling the semantics of spatial relations between named hydrologic features. The goal was to build a prototype application which can query a semantically enabled gazetteer of hydrological features of different feature types.

The GazOntology discussed in Chapter 3 presented the gazetteer ontology, which conforms with ISO standards to represent geographic information. The developed ontology showed that a feature can be identified with a spatial reference. The SpatialReference class is a generalization of two types of identifiers: (1) GeoIdentifier which can be the Official Name or an Alternate Name, and (2) Geometry which may be a point, polyline or polygon. Each hydrological feature can be identified by a name and geographic location. In addition to this, location equivalency is established between the different spatial footprints and feature names belonging to the same individual feature. The GazOntology was developed as an OWL ontology with OWL and RDFS properties.

Chapter 4 discussed the HydroOntology that models the canonical forms of real world hydrological features and their relationships. The developed HydroOntology can be aligned with DOLCE foundational ontologies as Physical Endurants. All the hydrological features can be seen as specializations of physical endurants.

Chapter 5 identified the canonical forms of hydrological features as represented in the NHD Dataset. The NHD Data model which is stored as a native ArcGIS geodatabase contains the hydrological features as NHDPoint, NHDFlowline and NHDWaterbody datasets along with Watershed Boundary Dataset. Topological relations between Stream-Stream, Stream-Lake, Lake-Lake were identified and placed in the semantic context of hydrology. Various spatial analysis methods and scripts were adapted and developed to extract hydrological relationships as modelled in the HydroOntology. RDF statements which describe the instances of classes and properties modelled in the HydroOntology were generated. Topological relations identified by the 9^+-intersection model between directed lines and regions are placed in the hydrological context.

Chapter 6 discussed the actual implementation of the HydroGazetteer prototype. The gazetteer of hydrological features was created as a triple store in Allegrograph. The GazOntology and HydroOntology were imported into Allegrograph to allow reasoning on the HydroGazetteer. A simple web application allows users to select hydro relations and search with a feature name. Predefined SPARQL queries are substituted with the entered query name and posted to the Allegrograph triple store as XML HTTP Requests. The response in JSON format is processed to display results on the map whenever applicable along with the tabular display of results. We demonstrated that it is possible to search a gazetteer based on the modelled relationships between named features. 7.2 Major Results and Discussion

Ontologies that support the semantic enablement of a gazetteer were developed. The Gazetteer Ontology (GazOntology) and Hydrological feature Ontology (HydroOntology) formed the ontological schema for the HydroGazetteer. These ontologies provided a set of classes and relationships in order to capture the topological relations between hydrological features in natural language terms that are specific to the hydrological domain.

The National Hydrography Dataset was identified as the data source to extract the canonical features and relations defined in the developed ontologies. Spatial analysis methods to identify key feature-feature relations such as Source, Mouth, Tributaries, MainStem, Upstream, and Downstream features, and flow relations, such as inflow, outflow, flows into, flows from, and flows through were developed using ESRI's ArcGIS product suite and Python scripting language. A total of 18 python scripts were developed and these scripts output RDF statements describing a hydrological feature by its name, identifier, spatial reference and spatial footprint of the feature and its relationships with other hydrological features. Topological relations between features as identified by the 9-intersection were evaluated as feature-to-feature relationships.

The HydroGazetteer was populated with 9251 instances of hydrological features (4759 streams, 3293 lakes, 85 rapids, 203 dams, 255 freshwaterbays, 656 saltwaterbays) and their hydrorelations (29532 stream-stream relations, 7185 stream-lake relations, 255 lake-freshwaterbay relations, 584 saltwaterbay-saltwaterbay relations) along with their spatial footprints were stored in 197408 statements in Allegrograph Triple store. Four groups of competency questions were used to evaluate the developed triple store. A simple user interface to select a hydrological relationship and a hydrological feature name was developed and the results were displayed in a USGS topographic base map. SPARQL queries were formulated to demonstrate the results for each of the four groups of competency questions. The implementation was able to demonstrate appropriate responses to competency questions such as:

- By searching for “BaysOf Frenchman Bay” all the bay names that are contained within Frenchman Bay were retrieved. This query demonstrates that spatial semantics can be used to improve the completeness of place name search results.
- Feature-feature semantic relationships map to other relationships (e.g., Topological and mereological relations), such as a search for “Tributaries of Mopang Stream” returns all tributaries thatare connected to the main stem of the Mopang Stream and are also a hydrological part of the river system.
- A spatial search for Machias River with Machias Bay as mouth excluded the Machias River which is a tributary of Aroostook River. This demonstrates that spatial semantics can support query disambiguation and complex spatial analysis such as upstream, downstream navigation is possible with place name searches.

In the light of these findings, the prototype implementation supports the original hypothesis: “*Semantic Feature-Feature relationships are derivable from Spatial geometry relations subject to domain constraints.”*

7.2 Future Work

The thesis demonstrated the ability to query on semantic relations among feature types at the level of named features. Further refinements of the approach are possible by considering parts of features in more detail. In this work the geometry of the feature is connected directly to the geoidentifier of the feature. This approach makes sense when the entire feature is included for a search result. However, streams are divided into reaches for hydrological analysis purposes and it is common to include the data for a specific reach within a River X. If the geometry is connected to the feature with the reach code of the stream segment, more meaningful results will be possible. For example, searches like "Waterbody at the Point Of Confluence of two Rivers X,Y" will provide the water bodies in the vicinity of confluence of the two rivers, instead of all the lakes along the rivers.

The current scope of this work includes only named hydrological features and hence query results can leave out features which have missing places names in the NHD dataset. Future work may include addressing data gaps in the NHD dataset and improve the accuracy of the query results.

Future work could also consider extensions to include hydrological processes, such as precipitation, evaporation, infiltration as perdurant classes to represent hydrologic processes and events. Participation relationship can be established between endurant and perdurant classes to model events that lead to a spatial change in a given time frame. Endurants and perdurants can be connected with a temporal index.

River impactedBy<Precipitation event>

PrecipitationEvent hasTimeIndex T1

Perdurant events can be mapped to endurants located in a geographic extent and spatio-temporal patterns in events can be studied and represented. Such a temporal extension could answer questions such as *What are the downstream features affected by a PrecipitationEvent near River X? or within watershed Y.*

BIBLIOGRAPHY

Berners-Lee, T., & Cailliau, R. (1994). Henrik Frystyk Nielsen, and Arthur Secret. *Communications of the ACM 37.8. The world-wide web., 37.8* , 76-82.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american 284.5*, 28-37.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International journal on semantic web and information systems 5.3* , 1-22.

Brank, J., Grobelnik, M., & Mladenić., D. (2005). A survey of ontology evaluation techniques.

Buchel, O., & Hill, L. L. (2011). Treatment of Georeferencing in Knowledge : North Americal Contributions to Integrated Georeferencing. *NASKO 2.1* , 47-57.

Buscaldi, D. (2011). Approaches to disambiguating toponyms. *SIGSPATIAL Special 3.2* , 16-19.

Chow, V., Maidment, D., Mays, L., Chow, V., Maidment, D., & Mays, L. (1988). *Applied Hydrology*. New York: McGraw-Hill.

Clementini, E., Felice, P. D., & Oosterom., P. v. (1993). A small set of formal topological relationships suitable for end-user interaction. *Advances in Spatial Databases. Springer Berlin Heidelberg*, 277-295.

Cohn, A. G., Bennett, B., Gooday, J. M., & Gotts, N. (1997). RCC: a calculus for region based qualitative spatial reasoning. *GeoInformatica, 1(3),*, 275-316.

Devaraju, A., & Kuhn, W. (2010). A Process-Centric Ontological Approach for Integrating Geo-Sensor Data. *FOIS*, 199-212.

Dingman, S. (1994). *Physical Hydrology* . Englewood Cliffs: Prentice Hall.

Dolbear, C., Hart, G., & Goodwin, J. (2006). What OWL Has Done for Geography and Why We Don't Need it to Map Read. *OWLED.*

Egenhofer, M. (1989, June). A formal definition for Binary topological relationships. *Lecture notes in Computer Science, 367*, 457-472.

Egenhofer, M. (2002). Toward the semantic geospatial web. *Proceedings of the 10th ACM international symposium on Advances in geographic information systems*, (pp. 1-4).

Egenhofer, M. J., & Franzosa, R. D. (1991). Point-set topological spatial relations. *International Journal of Geographical Information System*(5.2), 161-174.

Egenhofer, M. J., & Herring, J. (1990). Categorizing binary topological relations between regions, lines, and points in geographic databases. *The 9*, 94-1.

Egenhofer, M. J., & Herring, J. (1991). Categorizing binary topological relations between regions, lines, and points in geographic databases. *The 9*, 95-117.

Egenhofer, M. J., Sharma, J., & Mark, D. M. (1993). A critical comparison of the 4-intersection and 9-intersection models for spatial relations: formal analysis. *AUTOCARTO-CONFERENCE-. ASPRS AMERICAN SOCIETY FOR PHOTOGRAMMETRY*.

Fu, G., Jones, C. B., & Abdelmoty, A. I. (2005). Building a Geographical Ontology for Intelligent Spatial Search on the Web. *Databases and Applications.*

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L. (2002). Sweetening Ontologies with DOLCE. *Knowledge engineering and knowledge management: Ontologies and the semantic Web*, 166-181.

Goodchild, M. F. (2000). Communicating geographic information in a digital age. *Annals of the Association of American Geographers 90.2* , 344-355.

Goodchild, M. F., & Hill, L. L. (2008). Introduction to digital gazetteer research. *International Journal of Geographical Information Science 22.10* , 1039-1044.

Goodwin, J., Dolbear, C., & Hart, G. (2008). Geographical linked data: The administrative geography of great britain on the semantic web. *Transactions in GIS 12.s1*, 19-30.

Grenon, P., & Smith, B. (2004). "SNAP and SPAN: Towards dynamic spatial ontology.". *Spatial cognition and computation 4.1*, 69-104.

Guarino, N. (1997). Semantic matching: Formal ontological distinctions for information organization, extraction, and integration. *Information Extraction A Multidisciplinary Approach to an Emerging Information Technology*, 139-170.

Guarino, N. (1998, June 6-8). Formal ontology in information systems. *Proceedings of the first international conference (FOIS'98), 106*.

Hastings, J. T. (2008). Automated conflation of digital gazetteer data. *" International Journal of Geographical Information Science 22.10*, 1109-1127.

Herring, J. R. (1991). The mathematical modeling of spatial and non-spatial information in geographic information systems. *Cognitive and linguistic aspects of geographic space.*, 313-350.

Hill, L. L. (2000). Core elements of digital gazetteers: placenames, categories, and footprints. *Research and advanced technology for digital libraries*, *Springer Berlin Heidelberg*, 280-290.

Hill, L. L. (2009). Georeferencing: The geographic associations of information.

Hill, L. L., & Zheng, Q. (1998). Indirect geospatial referencing through place names in the digital library: Alexandria digital library experience with the developing and implementing gazetteers: Analysis and preliminary evaluation of the classical digital library model. *Proceedings of the Annual Meeting-American Society for Information Science. Vol. 36. Information Today,* 57-69.

http://www.geonames.org/about.html. (n.d.). Retrieved from http://www.geonames.org/about.html

http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/. (n.d.).

http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/. (n.d.). Retrieved from http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/

http://www.w3.org/TR/sparql11-query/. (n.d.).

ISO. (2002). Geographic Information - Spatial Referencing by Coordinates. *ISO/FDIS 19111:2002(E).*

ISO. (2003). Geographic Information - Spatial referencing by geographic identifiers. *ISO 19112:2003(E).*

Jackson, S. R. (2014). *RiverML: a harmonized transfer language for river hydraulic models.* PhD diss.

Janowicz, K., & Keßler, C. (2008). The role of ontology in improving gazetteer interaction. *International Journal of Geographical Information Science 22.10*, 1129-1157.

Jones, C. B., Abdelmoty, A. I., & Fu., G. (2003). Maintaining ontologies for geographical information retrieval on the web.On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. 934-951.

Jones, C. B., Alani, H., & Tudhope, D. (2001). Geographical information retrieval with ontologies of place. *Spatial information theory. Springer*, 322-335.

Klien, E., & Lutz, M. (2005). The role of spatial relations in automating the semantic annotation of geodata. *Spatial Information Theory*, 133-148.

Kurata, Y., & Egenhofer, M. J. (2006). The head-body-tail intersection for spatial relations between directed line segments. *Geographic Information Science.*, 269-286.

Kurata, Y., & Egenhofer, M. J. (2006). Topological relations of arrow symbols in complex diagrams. *Diagrammatic Representation and Inference.Springer Berlin Heidelberg*, 112-126.

Kurata, Y., & Egenhofer, M. J. (2007). The 9+-Intersection for Topological Relations between a Directed Line Segment and a Region. *BMI*, 62-76.

Manola, F., Miller, E., & McBride, B. (2001). RDF primer. *W3C recommendation 10*, 1-107.

Mark, D. M. (1999). Spatial representation: a cognitive view. *Geographical information systems: principles and applications*, 81-89.

Mark, D. M., & Egenhofer, M. J. (1994). "Calibrating the meanings of spatial predicates from natural language: Line-region relations. *Proceedings, Spatial Data Handling 1994., 1*.

Mark, D. M., & Egenhofer, M. J. (1994). Modeling spatial relations between lines and regions: combining formal mathematical models and human subjects testing. *Cartography and geographic information systems*(21.4), 195-212.

McKay, L., Bondelid, T., Dewald, T., Rea, A., Johnston, C., & Moore, R. (2012). NHDPlus Version 2: User Guide.

Obrst, L., Ceusters, W., Mani, I., Ray, S., & Smith, B. (2007). The evaluation of ontologies. *In Semantic Web. Springer US.*, 139-158.

Overell, S. (2011). The problem of place name ambiguity. *SIGSPATIAL Special 3.2*, 12-15.

Pequet, D., & Ci-Xiang, Z. (1992). An algorithm to determine directional relationship between arbitrary shaped polygons in the plane. *Pattern Recognition, 20.1*, 65-74.

Perry, M., & Herring, J. (2011). *"OGC GeoSPARQL-A geographic query language for RDF data. OGC Implementation Standard.*

Perry, M., & Herring, J. (2012). *OGC GeoSPARQL - A Geographic Query Language for RDF Data.* Open Geospatial Consortium project document OGC 11-052r4, v. 1.0. .

Peters, N. E. (1994). Hydrological Processes. In B. Moldan, & J. Cerny, *Biogeochemistry of Small Catchments: A Tool for Environmental Research* (pp. 207-228). John Wiley & Sons Ltd.

Peuquet, D. (1986). The use of spatial relationships to aid spatial database retrieval. *Proceedings of the 2nd International Symposium on Spatial Data Handling (SDH),* 459-471.

Prud'Hommeaux, E., & Seaborne, A. (2008). SPARQL query language for RDF. *W3C recommendation 15*. Retrieved from http://www.w3.org/TR/rdf-sparql-query/

Pullar, D., & Egenhofer, M. (1988). Toward formal definitions of topological relations among spatial objects. *Proceedings of the Third International Symposium on Spatial Data Handling, 225*.

Randell, D. A., Cui, Z., & Cohn, A. G. (1992). A spatial logic based on regions and connection. *KR 92*, 162-176.

Schwering, A. (2004). Semantic Neighborhoods for Spatial Relations. *Third International Conference on Geographic Information Science (GIScience)*.

Schwering, A., & Raubal, M. (2005). Spatial relations for semantic similarity measurement. *Springer Berlin Heidelberg*, 259-269.

Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The Semantic Web Revisited. *IEEE Computer Society*, 1541-1672.

Shariff, A. R., Egenhofer, M. J., & Mark, D. M. (1998). Natural-language spatial relations between linear and areal objects: the topology and metric of English-language terms. *International journal of geographical information science 12.3*, 215-245.

Sinha, G., Mark, D., Kolas, D., Varanka, D., Romero, B. E., Feng, C.-C., . . . Sorokine, A. (2014). An Ontology Design Pattern for Surface Water Features. *Geographic Information Science*, 187-203.

Smith, B. (1998). Basic Concepts of Formal Ontology.

Stadler, C., Lehmann, J., Höffner, K., & Auer, S. (2012). Linkedgeodata: A core for a web of spatial open data. *Semantic Web, 3(4)*, 333-354.

Taylor, P. (2012). *OGC® WaterML 2.0: Part 1- Timeseries.* Open Geospatial Consortium Inc.

Usery, L. E., & Varanka, D. (2012). Design and development of linked data from the national map. *Semantic Web 3.4*, 371-384.

Vilches-Blázquez, & Luis M., e. a. (2010). GeoLinked data and INSPIRE through an application case. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM*.

Wiegand, N. (2010). Ontologies and database management system technology for The National Map. *Cartographica: The International Journal for Geographic Information and Geovisualization 45.2*, 121-126.

Winter, T. C. (2001). Water source to four US wetlands: implications for wetland management. *Wetlands 21.4*, 462-473.

Wu, Y., & Winter, S. (2009). Inferring relevant gazetteer instances to a placename. *10th International Conference on GeoComputation*.

BIOGRAPHY OF THE AUTHOR

Nagalakshmy (Naga) Vijayasankaran was born on 20th May 1982, in the city of Pallayankottai, Tamil Nadu, India. She completed high school at St. Joseph’s Matriculation School, Madurai, India in 1999. She enrolled for her engineering degree at College of Engineering, Guindy at Chennai, India and graduated with Bachelors in Engineering (GeoInformatics) and Bachelors in Information Technology in a span of 5 years (1999-2004). She has previously worked for EDS and Telvent for a span of 4 years in various capacities as GIS Developer and GIS Analyst and worked on projects involving ESRI technologies. Naga Vijayasankaran is a candidate for the Master of Science degree in Spatial Information Science and Engineering from The University of Maine in May 2015.