Thesis Report

The semantic structure of spatial questions in human geography

Joris Wieleman

04-06-2018

Supervisor: Simon Scheider
Utrecht University
Professor: Stan Geertman
Utrecht University
j.g.wieleman@students.uu.nl

Contents

List of abbreviations

ADA – Analysis Data

CCD – Core Concept Data

GIS – Geographic Information System

OWL – Web Ontology Language

RDF – Resource Description Framework

RDFS – RDF Schema

SPARQL – SPARQL Protocol And RDF Query Language

SQO – Spatial Question Ontology

W3C – World Wide Web Consortium

XML – Extensible Markup Language

List of Figures

List of tables

List of Listings

Abstract

Due to a growing demand for spatial computing and an unlimited amount of possible spatial questions, there is a need to find common structures that repeat among questions in order to handle the variety. This would allow analysts to be better able to find the most optimal GIS tools and data. Consequently, this research project has investigated the semantic structure of spatial questions in Human Geography. The key to understanding the semantic structure of spatial questions lies in understanding the relations between the different pieces of information which are contained within a question. These relations are represented in a new lightweight ontology, the Spatial Question Ontology (SQO). The main purpose of SQO is to enable the description of all the information contained within spatial questions, which is necessary to answer the question, in a machine-readable format. Key elements which are covered by the ontology include: spatial data structure, question types, the spatiotemporal limitations of a question and core concepts of spatial information (e.g. object, field, event), which enable analysts to answer different kinds of spatial questions. In order to validate the ontology, a collection of spatial question from human geographic scientific literature was formalized into a Linked Data format. Geospatial ontologies such as SQO can be used for automated reasoning and could possibly enable a machine to find out whether a resource is applicable for a certain geo-analytical task. This may ultimately form the basis for question-based tool interaction and data analysis.

1. Context and background

1.1 Problem context and background

Data scientists in the field of Human Geography are currently using a wide range of different Geographic Information System (GIS) tools and related data to answer spatial questions. Due to a growing demand for spatial computing in business and government operations, GIS tools have been developed into vast and complicated toolboxes which require extensive technical knowledge to be applied efficiently. This process of increasing functionality can be described as 'creeping featurism', where incremental extensions to existing tools are favoured instead of a broad approach in order to design the next generation of GIS tools (Vahedi, Kuhn, & Ballatore, 2016).

When compared to other types of systems, a specific property of a GIS is the capability to pose questions and receive answers in a spatial context. However, existing user interfaces of GIS software do not directly support this (Gao & Goochild, 2013). Also, the growing complexity of GIS tools can make it difficult for users to answer spatial questions, especially for users from other domains without in-depth GIS skills. For these users without comprehensive knowledge about GIS, it is hard to identify specific functionality within one of the tools (Gao & Goochild, 2013).

The type of geo-analytical question often determines which tool(s) and data can be used to provide an appropriate answer. However, given the available variety of tools, selecting a suitable tool can be quite a tedious process. Because not every GIS user has the knowledge or experience which is necessary to link a certain spatial question to the corresponding GIS functionality (Gao & Goochild, 2013). Also, as the amount of available data is always increasing, this makes the search of finding the most suitable dataset(s) harder as well. In order for analysts to be better able to find the most optimal GIS tools and data, it is therefore necessary to investigate the different types of spatial questions which are asked and can be answered with a GIS tool. Given the fact that there is an unlimited amount of spatial questions that can be posed, there is a need to find common structures that repeat among questions as well as answers in order to handle this variety. Consequently, the focus of this research project will lie on investigating the semantic structure of spatial questions in Human Geography.

In this research, the following definition of a spatial question will be applied:

A spatial question is a question where the concept of location is necessary for formulating the answer.

As noted by Kuhn & Ballatore (2015), the concept of location is the most fundamental concept of spatial information, since it is always linked to location in some way. Besides location, Kuhn & Ballatore (2015) defined a list of other concepts which are core to spatial information. This includes concepts such as object, field and event, which are deemed to be an important aspect of the semantic structure of spatial questions.

Geospatial ontologies can be used for automated reasoning, and when done correctly, could possibly enable a machine to find out whether a Web resource is in fact suitable for a certain geo-analytical task (Scheider & Tomko, 2016). This may ultimately form the basis for question-based tool interaction and data analysis. Therefore, this research will attempt to apply Semantic Web technologies in order to develop an ontology which enables the translation of spatial question in human geography into a machine-readable format.

1.2. Research Objectives

This research project is focused on investigating the semantic structure of spatial questions in Human Geography. The main research objective was to empirically investigate spatial questions and the involved spatial concepts in order to develop an ontology which describes geo-analytical questions. This will ultimately stimulate a wider adoption of spatial analysis and mapping across various disciplines. Furthermore, this research might form the basis for future question-based tool interaction and data analysis.

1.3 Research questions

Resulting from the research objectives, the following main research question can be defined:

What is the semantic structure of spatial questions in human geography?

In order to answer this main research question according to the research objectives mentioned above, three sub-questions have been formulated:

1. *What is the variety of spatial questions asked in human geography?*

The purpose of this question is to investigate spatial questions in human geography and provide a set of linked data containing spatial questions collected from scientific literature. The creation of this dataset has allowed for the empirical investigation of spatial questions in human geography.

2. *Which spatial concepts are relevant for spatial questions in human geography?*

Core concepts of spatial information enable analysts to formulate and answer different kinds of spatial questions. Therefore, it is important to examine which of those are important in the field of human geography. The main goal of this question is to determine which spatial concepts, such as the concepts of spatial information mentioned in the work of Kuhn (2012) and (Kuhn & Ballatore, 2015), are relevant for spatial questions in human geography. This has provided the background knowledge necessary to be able to develop an ontology for relevant spatial concepts.

3. *How can an ontology be developed which covers the semantic structure of spatial questions in human geography?*

The final research question has combined the findings of the first research questions in order to develop an ontology which covers the investigated subject: spatial questions in human geography. Consequently, it has been validated whether the set of questions collected for the first research question was expressible by this ontology. Also, a statistical analysis based upon the set of linked data containing spatial questions has been provided. This also includes a reflection on the data and GIS tools which are necessary to answer these questions.

1.4 Research limitations

This research was focused on investigating the semantic structure of the spatial questions and the involved spatial concepts. Hence, this research is not about the actual development of an interface that captures spatial questions and identifies the required GIS functionalities based upon the spatial purpose of the question. Even though that would be a logical and compelling next step in research concerning this matter, it is beyond the scope of the current project. Furthermore, studying all the available human geographical studies is not feasible for the purpose of this research project. Thus, the focus was to provide a cross section of human geography studies, but it was not be possible to provide a comprehensive literature review.

1.5 Relevance

Developing an understanding of the semantic and syntactic structure of spatial questions has the potential to form the basis for future research on question-based tool interaction and data analysis. By combining and building on current research regarding core concepts of spatial information (Kuhn, 2012; Kuhn & Ballatore, 2015; Vahedi et al., 2016), as well as work about research questions (Jarvinen, 2004) and Semantic Web technologies (Heath & Bizer, 2011; Perez, Arenas, & Gutierrez, 2006), this current research provides a way to formalize spatial questions in order to make them machine readable.

The societal relevance of this research project stems from the fact that by developing an ontology for spatial concepts and questions, this may open the door for applications which offer a more efficient search method to look for appropriate spatial analytical tools and suitable data. Ultimately, it could stimulate a wider adoption of spatial analysis and mapping in Human Geography, as well as across various other disciplines where spatial questions play a role.

2. Theoretical Background

The aim of this chapter is to place the current research within its scientific context. Relevant and similar studies will be discussed as well as how the knowledge gained from these can be applied in the current research. The chapter is subdivided based on some of the most relevant concepts for this study: spatial concepts, spatial research questions and the semantic web technologies.

2.1 Spatial concepts

A distinctive property of spatial questions is that they can be formulated using spatial concepts (Kuhn, 2012; Scheider, Gräler, Pebesma, & Stasch, 2016).There is a growing number of projects which suggest that the development of a conceptual view of spatial information can greatly benefit the understanding of environmental, cognitive and social processes (Kuhn, 2012). In order to be able to investigate spatial questions, it is therefore necessary to examine the spatial concepts involved. Currently, the research regarding what the most fundamental spatial concepts are is an ongoing process. Therefore, recent progress on the matter will need to be examined.

In his original work on the matter, Kuhn (2012) describes ten core concepts of spatial information which can be related to geo-analytical questions. This earlier set of core concepts and included six spatial concepts and four information concepts can be found in table 2.1. The distinction was made between spatial concepts, which describe space, and information concepts, which describe spatial information. For example, the spatial concept of location refers to space and is used to interpret spatial data. However, information concepts do not necessarily have to be spatial. An example is the information concept value, this can refer to spatial information without actually being spatial. Granularity is an example of an information concept which is spatial, since it is a spatial measure as well as it is used to interpret spatial data.

Table 2.1: Core concepts of spatial information proposed by Kuhn (2012)

Spatial concepts	Location
	Field
	Object
	Network
	Event
	Neighborhood
Information concepts	Granularity
	Accuracy
	Meaning
	Value

In a more recent article however, Kuhn & Ballatore (2015) proposed a smaller set which contains seven spatial concepts. Here, the categorical distinction between concepts was made based on information content and information quality. The most recent version of this set can be found in table 2.2, which was updated in Kuhn (2017), and now contains one base concept, four content concepts and two quality concepts. Originally, location was considered as core content concept as well but was regarded as the base concept of spatial information in Kuhn (2017) since it is now considered to be the most fundamental concept. Neighborhood as a spatial concept was removed from the list since it cuts across the other content concepts, as well as to the notion of location. Furthermore, value and meaning where removed since there are currently no solid theories which make them 'core' to spatial analysis. As noted in Kuhn & Ballatore (2015), this list may still miss some ideas regarding spatial concepts or relegates them to sub concepts or non-cores status.

Table 2.2: Core concepts of spatial information proposed by Kuhn & Ballatore (2015)

Base concept	Location
Content concepts	Field
	Object
	Network
	Event
Quality concepts	Granularity
	Accuracy

Besides the work of Kuhn (2012) and Kuhn & Ballatore (2015), other attempts of capturing spatial thinking in a small set of concepts without disciplinary, mathematical, or technological biases where also made. Examples include the list of 'foundation concepts in spatial thinking' put forward by Jannelle & Goodchild (2004). However, as noted by Kuhn (2012), this list strongly focuses on spatial thinking and the social sciences instead of spatial information and its properties in general.

Based upon the summary of previous research regarding the subject of spatial concepts, the choice was made to further examine the core concepts of spatial information from Kuhn (2017). The following part of this section will therefore aim to provide further insight into these seven concepts and discussing their properties.

2.1.1 Location

Spatial information is always linked to location in some way and is therefore considered to be the most fundamental concept of spatial information. The base concept of location has the primary purpose of asking and answering where-questions (Kuhn, 2017; Kuhn & Ballatore, 2015). An important notion is that location is not a property or attribute, but a relation. This is due to the fact that locating something relies on spatially relating it to something else (Donnelly, 2005). Every location description expresses *spatial relations* between *figures* to be located and chosen *grounds* (Kuhn, 2012; Kuhn & Ballatore, 2015). In this relation, the thing that gets located is the figure and the thing that locates it is the ground(Kuhn & Ballatore, 2015). For example, in the location description 'The city of Utrecht lies within the Netherlands' Utrecht forms the figure and Netherlands is the ground. This implicates that information about location is always based upon some form of human judgment, since the assignment of figure and ground roles and relating them spatially inherently contains a choice made by people.

When considering a location, a distinction can be made between a *place* and a *position.* Places are often used as grounds and frequently have a name (e.g. The Netherlands, Utrecht or the North Sea). Places are, just as all grounds, spatiotemporal entities. This means that a certain place may eventually stop to fulfil its function as a ground (Kuhn & Ballatore, 2015). This is illustrated by places such as Czechoslovakia and East Germany, both are countries which ceased to exist during the previous century. Place-based location information is a common aspect of human communication and has also become a key resource in spatial computing (Kuhn & Ballatore, 2015). Positions describe spatial relationships in a quantitative manner by using distances from the grounds established by coordinate systems. This is also the case in the vertical dimension. For example, heights use a geoid or ellipsoid as ground and distances from it as position (Kuhn & Ballatore, 2015).

2.1.2 Field

Field information allows for the answering of questions about the value of an attribute anywhere in a space of interest. This is because fields describe phenomena which can be described by a property with a single value at any given position in the space of interest (Kuhn & Ballatore, 2015). An example of a field could be the air quality in Utrecht. When examined mathematically, all fields are characterized by a continuous function from positions to values. This means that a small change in position amounts to a small change in value. These positions can be spatiotemporal, however, time is often detached from space and modelled as a snapshot. Also, both the positions and values can be discrete while still allowing for continuous functions between them (Kuhn & Ballatore, 2015). In reality however, the condition of continuity on the function is ignored occasionally, which results in keeping only the functional relationship between positions and values (e.g. the phenomenon of land cover). The concept of a field is one of the two fundamental views of spatial information, the other being an object (Kuhn & Ballatore, 2015).

2.1.3 Object

As the second fundamental view of spatial information, objects capture individual things which are extended in space that can be identified and described by relations and properties. Object information can therefore answer questions about the properties and relations of objects (Kuhn & Ballatore, 2015). All kinds of objects have only one defining characteristic, which is that they hold an identity. The purpose of this is that the properties and relations of an object can be tracked over time. All other properties and relations of an object can change. This is also true for spatial relations, as objects can be mobile as well (Kuhn & Ballatore, 2015). Furthermore, all objects are bounded, which means that they all have a finite size, even though their exact boundaries may not always be known. Beaches, for example, are objects with ambiguous boundaries which can better be described by using a transition zone between what clearly is part of the object and what is not. Another important notion for objects is that their properties and relations apply to them as a whole and can be either spatial or thematic. (Kuhn & Ballatore, 2015). Object types can be determined based upon shared properties and relations. Since there are a lot different types of objects, it may also be useful to use classifications (Kuhn & Ballatore, 2015).

2.1.4 Network

When there are connections between objects, this information is stored in networks. Networks can be applied to answer various questions about connectivity (e.g. what is the fastest route?). Most of the time, networks are modelled as graphs and this results in the availability of a large number of algorithms and implementations associated with this. Nodes are objects connected in a network and the array of objects that can be nodes is unlimited. Edges in a network are formed by any binary relation which connects a pair of nodes. These can have a physical presence in the world as well as be abstract connections, such as a social network (Kuhn & Ballatore, 2015).

There are various properties which can apply to networks. For example, the edges of a network may be directed. Also, a relevant property of an edge is often characterized by a single attribute. This can be either numeric or nominal attributes (Kuhn & Ballatore, 2015). In the case of numeric attributes (weight), this can for instance be the travel distance and when nominal attributes (colour) are concerned, this could for example be a road-type.

2.1.5 Event

Information about events helps answer questions about change (Kuhn, 2012). This is swiftly becoming an important subject within the field of spatial information, fuelled by an increase in the availability of information about changes in a dynamic world (Kuhn & Ballatore, 2015). Events are separate parts of process and always have a finite duration, as well as an identity. Furthermore, an event is characterized by temporal and thematic properties and relations. The most common temporal relations which can occur between events are posteriority, precedence and co-occurrence. Similar to the previously mentioned ambiguous boundaries of objects, the start and end of an event may not always be clear (Kuhn & Ballatore, 2015). The most important relationship between the other core concepts of spatial information and events is participation. Events concern networks, objects and fields as participants and they are often changed because of the way they participate in a certain event (Kuhn & Ballatore, 2015).

2.1.6 Granularity

Questions regarding the level of detail in spatial information can be answered by the means of information about granularity. Essentially, all content concepts mentioned above are characterized by granularity (Kuhn & Ballatore, 2015). Granularity is a concept of information quality, because it tells something about information itself instead of other things on earth. Precision or resolution are similar terms which are sometimes used instead of granularity (Kuhn & Ballatore, 2015). Examples of granularity information include what the cell size is in a raster dataset or the length of rainfall during a storm.

The fact that temporal and spatial granularity are extents in time and space, make it possible to quantify granularity. However, because there are variations in the observation-to-visualization life cycle of information, it is quite difficult to decide on a measure for granularity which can be commonly applied (Kuhn & Ballatore, 2015).

2.1.7 Accuracy

Accuracy in terms of information is describing something correctly. Therefore, information about accuracy can provide answers to questions about the correctness of spatial information (Kuhn, 2012). Accuracy can only be established for a certain granularity and when regarding reference information that is thought to be accurate. By measuring the difference between this reference information and a certain piece of information, its accuracy can be determined. Locational accuracy is often focused upon; however, it also applies to all other information about fields, objects, networks and events. Together, granularity and accuracy are the most important indicators of the quality of spatial information (Kuhn & Ballatore, 2015).

Bias is one of the main notions associated with accuracy. This occurs if there is a difference between a hypothetical true value and the mean of repeated observations. However, when appropriate equipment for measuring is used, repeated measurements distribute regularly around the true value. This is the result of conceiving measurement as a random process. Also, accuracy is repeatedly related to the lack of systematic errors. Because of this, measurements are only affected by random errors. This can be achieved by calibration in order to diminish systematic errors to a level which is lower than the granularity of the data (Kuhn & Ballatore, 2015).

2.2 The nature of spatial research questions

In order to investigate the semantic structure of spatial questions, it is besides a deeper understanding of spatial concepts, also necessary to study the linguistic and syntactical nature of questions themselves. Defining the research question is one of the most important steps in the research process since these questions narrow down the objective and purpose of the research to specific questions that analysts attempt to address in their studies (Onwuegbuzie & Leech, 2006).

The type of (spatial) question often determines which GIS tools or research method can be used to come to an appropriate answer. This is supported by the work of Jarvinen (2004) and Onwuegbuzie & Leech (2006), who suggests that the research question is the essential factor when selecting a proper research method. In his article, Jarvinen (2004) outlines a taxonomy of six different research approaches which is shown in figure 2.1. This taxonomy follows a top-down approach, where a distinction is for example made between mathematical approaches and all research approaches which study reality. Other approaches are conceptual, analytical, theory-testing, theory creating, artifacts-building and artifacts-evaluating approaches. It can be argued that most, if not all, spatial questions in human geography answerable by a GIS fall in the category 'approaches for empirical studies', since these questions stress what reality. They are not part of the conceptual-analytical approach since spatial questions answerable by a GIS are ordinarily not part of method for actual theoretical development. Instead, these spatial questions are used as way to empirically study the past or present. Here, a differentiation can be made between theory-testing and theory-creating approaches. The difference between the two is based on the presence of a theory, model or framework which is guiding the research or whether a new theory is being created based upon gathered data (Jarvinen, 2004).

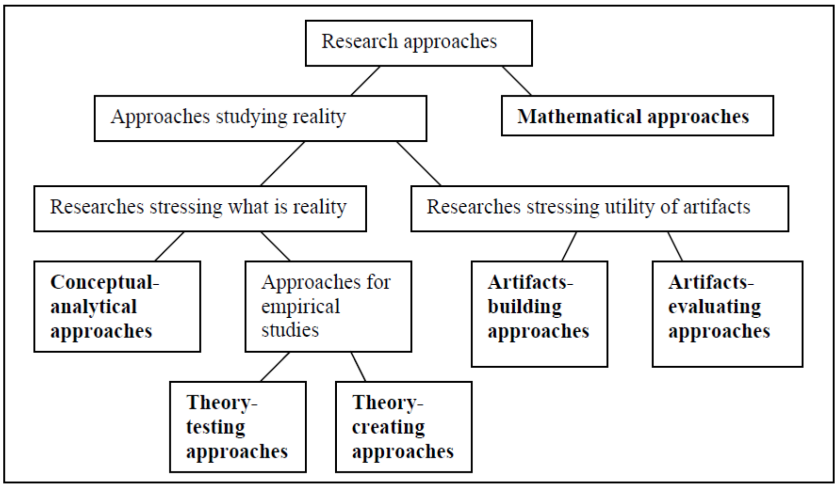

Figure 2.1: Taxonomy of research methods by Jarvinen (2004)

When considering a matter of interest, in this case spatial questions, both the aforementioned spatial concepts and the different types of research questions can be used in a collection of information known as an ontology (Maedche & Staab, 2001). Geospatial ontologies can be used for automated reasoning, and when done correctly, could possibly enable a machine to find out whether a Web resource is in fact suitable for a certain geo-analytical task (Scheider & Tomko, 2016). This may ultimately form the basis for question-based tool Interaction and data analysis. Therefore, it is necessary to examine the related Semantic Web technologies.

2.3 Semantic Web technologies

This section will discuss some of the most relevant Semantic Web Technologies which were applied in this research.

2.3.1 The Semantic Web

The World Wide Web has fundamentally changed the way people communicate and society as a whole. Currently, it is also at the basis of a revolution that is leading the developed world to be changed into a knowledge economy and, in broader terms, to a knowledge society (Antoniou & Van Harmelen, 2009). The Semantic Web originated as a vision for the future of the Web where information is given explicit meaning. At the time, the concept of the Semantic Web was often regarded as the next big technological leap for the Internet (Antoniou & Van Harmelen, 2009; Knublauch, Fergerson, Noy, & Musen, 2004). The Semantic Web can be seen as an extension of the existing Web which better enables machines to automatically process and integrate information available on the Web. Currently, the process of integrating Semantic Web technologies is well on its way (Antoniou & Van Harmelen, 2009; Berners-Lee, Hendler, & Lassila, 2001).

The Semantic Web makes use of a set of technologies known as the Semantic Web Stack. Figure 2.2 provides an overview of the Semantic Web Stack and is based on the architecture provided by Hoyland, Adams, Tolk, & Xu (2014). Outlined in red are the technologies which are most important to the current research: RDF, RDFS, OWL and SPARQL. These concepts will be addressed in the following paragraphs.

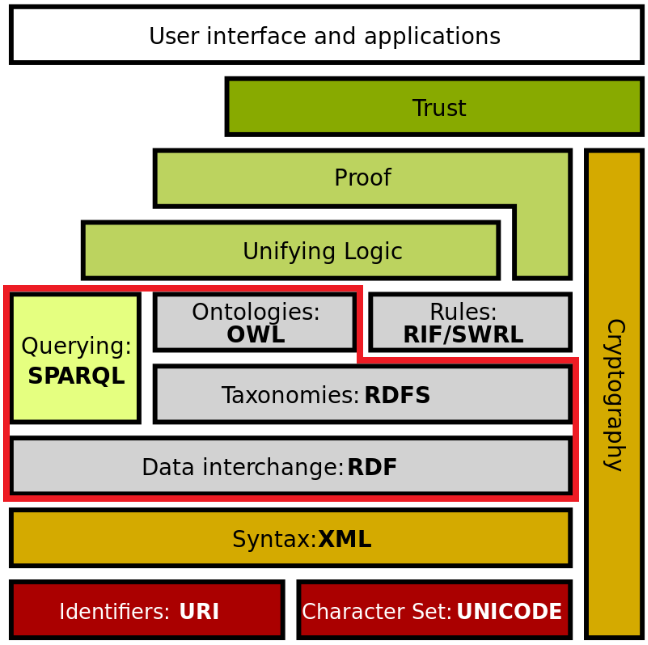

Figure 2.2: Semantic Web Stack

2.3.2 RDF

Even though the Resource Description Framework (RDF) is often described as a 'language', it is in fact a data model (Antoniou & Van Harmelen, 2009). RDF is the format for creating Linked Data for the Semantic Web. This labelled and directed graph data format is recommended by the World Wide Web Consortium (W3C) and is regularly used to represent social networks, personal information and also to facilitate integration of disparate information sources in the Semantic Web (Han et al., 2016). In RDF, the data is published in basic building blocks known as *triples.* Triples are statements which contain the three parts *subject, predicate* and *object* (Antoniou & Van Harmelen, 2009; Hoyland et al., 2014).

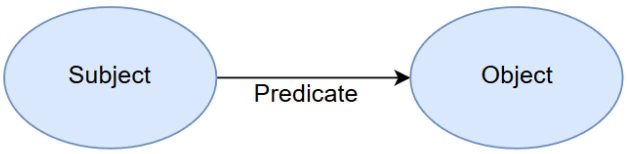

Figure 2.3: RDF triple statement

The subject-predicate-object structure of a triple is depicted in figure 2.3. In a triple, the subject is the resource which is being described, the predicate is a property of that resource and the object is the value of that property. Resources in RDF can be considered as an object or "something in the world" which we want to talk about (Antoniou & Van Harmelen, 2009; Cyganiak, Wood, & Lanthaler, 2014). Resources can be anything such as places, people, authors, or, in the case of the current research, journal articles which contains a spatial research in human geography. Properties are a special type of resource, since they describe the relations between resources. Each resource in RDF contains a Uniform Resource Identifier (URI), which can be a Uniform Resource Locator (URL, also known as a web address). However, an identifier does necessarily need to enable access to a certain resource (Antoniou & Van Harmelen, 2009; Cyganiak et al., 2014).

First, RDF was mostly represented by using XML syntax. However, there currently are a number of different serializations such as Turtle, N-Triples and JSON. Turtle is often preferred, as it provides the possibility to apply namespace prefixes, making it easier to read RDF triples and writing them. It is a plain text format (Heath & Bizer, 2011).

2.3.3 Information ontologies

In order to be able to structure data for comprehensive and transportable machine understanding, collections of information known as *ontologies* are required (Maedche & Staab, 2001). Ontologies are an important element of the Semantic Web because they provide formal models of domain knowledge that can be used by intelligent agents (Knublauch et al., 2004). Berners-Lee et al. (2006, p.28) describe the advantages of an ontology in the following terms:

"Ontologies contain specifications of the concepts that are needed to understand a domain […] a formality that makes the ontology machine-readable, and therefore allows a machine to undertake deeper reasoning over Web resources. The disadvantage is that such formal constructs are perceived to be hard to create."

The most common ontologies for the Semantic Web consist of a taxonomy, which formally defines classes of objects and relations among them, and a set of inference rules (Berners-Lee et al., 2001). This is also discussed in Noy & McGuinness (2001), where they outline several basic elements of an ontology. In their work, they describe an ontology as a formal explicit description of:

- *Concepts* within a domain (*classes*)
- *Properties* of each concept describing various features and attributes of the concept (also called *slots*)
- *Restrictions* on slots (also called *facets*)

When an ontology is combined with a set of individual *instances* of classes, a *knowledge base* is created. However, it is often hard to determine where the ontology ends and the knowledge base begins (Noy & McGuinness, 2001).

2.3.4 RDFS and OWL

A generic and abstract data model for describing resources is provided by RDF using triples. But this does not provide any domain-specific terms for describing classes. This is done through vocabularies, taxonomies and ontologies expressed in the Web Ontology Language (OWL) and RDF Schema (RDFS) (Heath & Bizer, 2011). RDF Schema (RDFS) is a language to describe lightweight RDF ontologies. RDFS generalizes the classes and properties of RDF resources through hierarchies. The most basic RDFS vocabularies contain class and property type definitions (Heath & Bizer, 2011).

OWL is designed for use by applications that need to process the content of information as opposed to presenting information in a human-readable format. Greater machine interpretability is facilitated by OWL by providing additional vocabulary along with a formal semantics. This is achieved through providing more modelling primitives (Heath & Bizer, 2011; McGuinness & Van Harmelen, 2004). Therefore, the Semantic Web may provide a way to formalize spatial questions in order to make them machine readable.

2.3.5 SPARQL

The SPARQL Protocol And RDF Query Language (SPARQL) is the standard language for querying RDF data (Perez et al., 2006). Since the release of RDF as a W3C recommendation in 1998, a number of designs and implementations of RDF query languages have been put forward. It was not until 2004 that a public working draft of SPARQL was first released. However, since then, it was rapidly adopted as the standard for querying data of the Semantic Web, resulting in becoming a W3C recommendation in 2008 (Antoniou & Van Harmelen, 2009; Perez et al., 2006).

In some way, SPARQL may be considered as the 'linked data equivalent' of SQL. However, instead of retrieving information from relational databases, it handles information from graph data models stored as triples. Essentially, SPARQL is a query language which is based on matching graph patterns (Antoniou & Van Harmelen, 2009; Perez et al., 2006). The triple pattern is the most basic graph pattern. This pattern resembles that of RDF triples, with the difference however that they can contain a variable instead of an RDF term in one of the subject, predicate or object positions. Combining triple patterns results in a basic graph pattern, after which an exact match to a graph is required in order to fulfil a pattern (Antoniou & Van Harmelen, 2009).

In the context of this research, RDF has provided the format for the creation of a collection of spatial questions in Human Geography. Consequently, this allowed for the SPARQL Protocol And RDF Query Language (SPARQL) to be applied on this dataset since this allows queries to consist of triple patterns.

3. Methodology

This chapter will outline the methodology for this research as well as describing the required resources and the main research outputs. First, a brief overview of the research is provided. The different research phases, in the form of a flowchart, can be found in figure 3.1. Here, the different steps which are necessary to answer the research questions are presented. The blue boxes indicate the individual steps of the project while the grey boxes order them per research phase. Secondly, the required software tools for this research are described. Hereafter, the individual research phases are discussed in more detail.

3.1 Research phases

The Exploration phase of the research broadly consisted of two distinct elements: the literature review conducted in the previous chapter and the creation of a collection of spatial questions published in scientific literature. This phase will be discussed in section 3.3

The Ontology Development phase is aimed at the development of the SQO. Developing competency questions, the ontology model and creating and extending SQO are the primary steps in this phase. Competency questions consist of a set of questions about a matter of interest, which are stated and replied in natural language. The goal of developing these questions is to identify the ontology requirements (Gangemi, 2005). Section 3.4 will cover this phase in more depth.

The Ontology Validation phase has the primary purpose of validating SQO. This is done by triplifying the dataset containing spatial questions from scientific literature created in the Exploration phase. Then, based on the input provided by the triplifying of the data and assessment of the competency questions, SQO has been revised or extended accordingly. Section 3.5 will go into more detail regarding this research phase.

The Analysis phase is the final phase of the research and has the purpose of analysing the RDF dataset containing the formalized spatial questions. This is done by creating several Python scripts to examine the dataset. The results will provide insight in the semantic structure of spatial questions in the field of human geographic scientific literature.

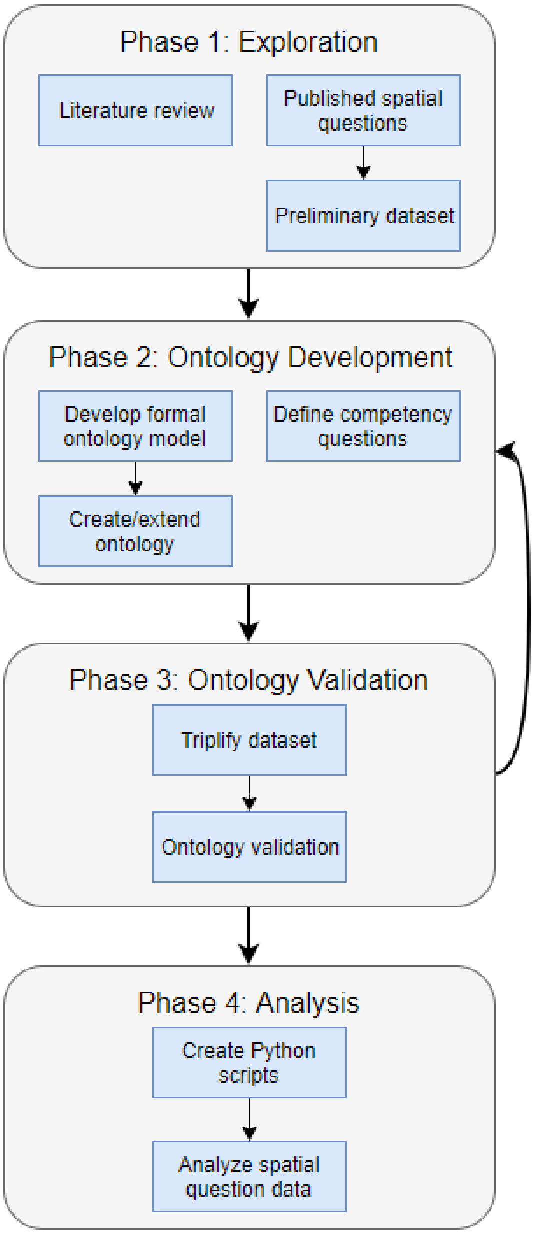

Figure 3.1: Research phases overview

3.2 Software tools

This research required several software tools which were necessary to develop and test SQO and analyse the results. This section focuses on three of the most significant tools which have been used in this process.

Protégé

Protégé, a free and open-source OWL editor, was used for the modelling of SQO. Protégé, developed by Stanford University, can be used to model new ontologies as well as extending existing ontologies. This includes all the necessary elements for ontology development such as defining classes, hierarchies and relations between the classes. Protégé can also be used to insert logical axioms in the ontology so that classes and relations can be inferred (Knublauch, Fergerson, Noy, & Musen, 2004; Knublauch et al., 2005). Furthermore, the possibility to import already published ontologies facilitates the extension of these ontologies.

DotNetRDFkit

For the creation of the RDF dataset, the DotNetRDF toolkit have been used, a suite of command line and graphical user interface (GUI) tools for working with RDF and SPARQL. This toolkit supports processes such as writing, searching, replacing and saving data to RDF files. Furthermore, the DotNetRDF library contains a query engine which is compatible with SPARQL. Also, it contains additional tools, like a file convertor, an editor which enables SPARQL and RDF editing and the use of a GUI (Miltiades & Angelides, 2016), making it a suitable tool for this research.

PyScripter

For the purpose of creating Python scripts to analyse the RDF dataset, PyScripter have been used. This software tool is a free and open source Python Integrated Development Environment (IDE). PyScripter is a lightweight but feature-rich IDE. It provides various additional functionalities when compared to other Python IDE's such as interactive syntax checking, creating projects, variable watch tools and window tabs (Tateosian, 2015).

3.3 Phase 1: Exploration

This section will outline how the Exploration phase will provide the necessary basis for the rest of the research to build upon. This phase is formed by the literature review, which was undertaken and discussed in chapter 2, and the collection of published spatial questions in scientific literature in the field of human geography. Since the literature review has already been discussed in the previous chapter, this section will solely focus on the data collection process.

3.3.1 Spatial question collection

Besides the insights gained from the literature review, the second main output of the Exploration Phase is to create a collection of spatial research questions from the field of human geography. The goal for this collection will be to contain 50 spatial questions which are found in recent scientific literature, as this is deemed a sufficient number of questions for validating the ontology and to provide meaningful statistics.

In order to create a collection containing a number of geo-analytical questions, a web search in scientific literature was conducted using the Scopus search engine. To do this, a search strategy was developed. There are several points which were taken into account when searching for scientific articles containing suitable spatial questions:

- That the articles included are about a subject in the field of Human Geography
- That the articles contain a spatial question and cover a GIS-related answer, which can be included in the RDF dataset
- That the articles are recent, written in the period from 2014-2018

Based upon these criteria, a search was executed in Scopus, since this is the largest abstract and citation database of peer-reviewed literature. The search query in listing 3.1 was used for this purpose:

```
(((human OR cultural OR economic OR feminist OR political OR urban)
AND geography) OR ((migration OR population) AND studies) OR
geopolitics OR travel OR tourism)
AND
(gis OR 'spatial analysis' OR 'geographical information' OR 'spatial
statistics')
```

Listing 3.1: Search query for scientific articles containing spatial questions from human geography

The first section of this query has the function to only select scientific literature which is about human geography. Therefore, the most common fields of study within human geography have been added to this part of the query. The second part of the query has the goal to make sure that there is some form of spatial analysis being conducted as part of the research within the article. Also, all articles outside the aforementioned period were excluded from the search.

The resulting number of hits from the above search query was 4764 entries of scientific literature. Based on this number, it was decided that every tenth piece of literature will be included in the research as part of the data if it indeed complies with all criteria. To ensure an even spread over the different years from this period, only ten spatial questions from each year were added to the dataset.

It was not possible to automatically extract the spatial questions from the articles; therefore, this has been done manually. For every included article it was checked whether it complied with the above criteria, and if so, added to the collection of questions. The full list of spatial questions included in the research can be found in Appendix A. Since it is quite rare for a scientific journal article or paper to contain a literal (spatial) research questions, it will be attempted to determine what the research question is based on the article**.**

Key places which were used to determine the spatial question are the title and abstract of an article. An example of this question extraction process is provided in figure 3.2, which shows the title and abstract of Qiao, Cao, Liu & Wu (2018). The red markings in the title indicate key sentences and phrases which helped extract the spatial question. In this case, the spatial question was '*How can the EPIC-model be applied to assess the simulation accuracies at different scales?'*. If it was not possible to extract a logical spatial question from the article, it was not be included in the collection of questions. An example of an article from which it was not possible to deduce a spatial question I provided in figure 3.3. Even though it is clear that this article by Fadhly & Matondang (2018) is about implementing what they call transit-oriented development in handling the congestion effect on urban sprawl phenomenon and traffic growth, the abstract does not provide a logical spatial question. Therefore, this article was not included in the collection of questions.

It must be noted that this is an arbitrary process, since it is possible that in some cases the extracted spatial question does not reflect the original intention of the authors. However, it is the only feasible method to extract a sufficient amount of geo-analytical research questions from the articles and this process was treated with the utmost care.

Scale dependence and parameter sensitivity of the EPIC model in the agro-pastoral transitional zone of north China (Article)

Qiao, J.[a], Cao, Q.[b], Liu, Y.[c], Wu, Q.[a]

[a]School of Geography and Environment, Shandong Normal University, Jinan, 250014, China
[b]Center for Human-Environment System Sustainability (CHESS), State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China
[c]Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, 361021, China

Abstract

View references (57)

Under the background of rapid population growth, achieving accurate assessments of crop yields is critical to national policy adjustment and food security. Among all the evaluation methods for crop yields, the Environmental Policy Integrated Climate (EPIC) model is considered to be an effective tool. Since published, it has been widely applied on local, regional, and global scales. However, in its practical applications, the impact of scale variation on the simulation results has not received full consideration. The scale variation can introduce many uncertain effects in the simulation results. Furthermore, as a basis for parameter calibration, parameters sensitivity analyses were always been ignored at large scale studies. In this study, we used the GIS-based EPIC (GEPIC) model to assess the simulation accuracies at different scales and local sensitivity analysis was adopted to analyze the sensitivity of the crop parameters in each grid first time. The results showed that EPIC model has scale-dependence and the smaller the scale is, the more accurate the results are. The most sensitive parameter of maize and soybeans is the energy conversion to biomass factor (WA) in the agro-pastoral transitional zone of north China, whereas the harvest index (HI) is the most sensitive parameter for the three other crops of rice, spring wheat, and winter wheat. For the same crop, parameter sensitivity is heterogeneous in space due to varied input variables. In addition, we find that parameters sensitivities (S_i) vary as the parameters and the variations in S_i for the five parameters are different. Excluding the S_i of HI for spring wheat which has the greatest change and the variation is nearly 0.5, the S_i of other parameters change slightly. In the application of GEPIC model, we should consider the effects of scale variation on crop yields, particular for areas with high heterogeneity of cropland patches. In addition, this study provides a reference for setting the input parameters of the EPIC model. © 2018 Elsevier B.V.

Figure 3.2: Title and abstract of Qiao, Cao, Liu & Wu (2018). Red markings indicate key phrases for question extraction.

Implementation of Transit Oriented Development in Handling Congestion Effect on Urban Sprawl Phenomenon and Traffic Growth in Banda Aceh

Fadhly, N.[a], Matondang, R.[b]

[a]Department of Civil Engineering, Engineering Faculty, Syiah Kuala University, Darussalam Banda Aceh, Indonesia
[b]Doctoral Program in Regional Planning, University of North Sumatra, Indonesia

Abstract

View references (15)

The occurrence of spatial and road network changes, increased traffic volume approaching road capacity, causing congestion due to urban sprawl phenomena as well as lack of public facilities and social facilities. From this background problem considered to be solved by a concept that is Transit Oriented Development (TOD), concept of transportation development in co-operation with spatial. The TOD concept can eliminate urban sprawl, which transforms urban sprawl into a compact city area. Pull model are obtained from the results of the analysis is, Oi = -37,126 + 58,723 X9 + 17,968X2, where road performance, land use, the coefficient indicates the number of additional trip generation. Trips pull obtained by model Dd = -20,351 + 30,903X2 where only one significant variable causing attraction of land use that is dominant. With concept of TOD, the model obtained is Oi = 32.180 + 0.002X7-7.017X3 where travel and distance travel costs indicate the number of additional trips generation, the generation will decrease over the base year. Objects obtained with the model Dd = 48.474-0.671X11 + 0, 003X7-9,299X3, meaning travel attraction in the absence of population density, travel expenses, and travel distance, then travel appeal that occurs on 48 journeys influenced by variable population density and distance travel will decrease, it can be interpreted this independent variables can influence the drag on the concept of Banda Aceh City and TOD can be applied in the city of Banda Aceh. © Published under licence by IOP Publishing Ltd.

Figure 3.3: Title and abstract of Fadhly & Matondang (2018).

3.4 Phase 2: Ontology Development

The main goal of this research is to develop and validate the ontology which can be used to describe the information contained within spatial questions. Hereby, the focus will be on information that is necessary to answer the question. This ontology has been created using Protégé, which is an open platform for ontology editing and knowledge acquisition in order to enable conceptual modelling with Semantic Web languages such as OWL (Knublauch et al., 2004; Noy et al., 2001). The OWL plugin, an extension for Protégé, allows the editing of OWL ontologies, acquisition of instances for semantic mark-up and the access of description logic (DL) reasoners. The choice to use Protégé and the OWL plugin was made because it offers several advantages: the benefits of a big user community, a flexible architecture and a library of reusable components (Knublauch et al., 2004).

SQO is a lightweight ontology, in this case consisting of a taxonomy and several associations between the classes. It can be seen as a directed graph of which the nodes represent concepts (classes) related to spatial questions. The links between the nodes indicate associations (properties) between the related concepts (Antoniou & Van Harmelen, 2004). It was developed as a lightweight ontology because it does not require any more expressiveness for its intended purpose, which is to describe the information contained within spatial questions.

3.4.1 Ontology engineering techniques

The lifecycle of ontologies involves a number of different techniques ranging from manual to automatic building, merging, refinement and mapping, among others. For this research, the manual technique currently seemed to be the most logical choice. However, all of these techniques require the specification of core concepts for the population of the ontology (Gangemi, 2005). In the current research, the spatial concepts examined in the previous chapter will play an important role in this process. In their work, Noy & McGuinness (2001) outline a number of ontology-design ideas which can be used to develop the ontology to describe spatial questions and spatial concepts in Protégé. In practical terms, the development of an ontology according to Noy & McGuinness (2001) includes the following:

- Defining the classes in the ontology
- Arranging the classes in a taxonomic (subclass–superclass) hierarchy
- Defining slots and describing allowed values for these slots
- Filling in the values for slots for instances

After these steps have been completed for the current research, a knowledge base can be created by defining individual instances of these classes filling in specific slot value information and additional slot restrictions (Noy & McGuinness, 2001).

The ontology which describes spatial questions and spatial concepts will then be used to translate the natural language spatial questions, gathered in the Exploration phase, into a structured form. Subsequently, this RDF dataset will allow for statistical analysis of the spatial questions and a way to test the quality and usability of the ontology created in the earlier phase of the research project.

3.4.2 Reused ontology sources

As mentioned by Noy & McGuinness (2001), reusing domain knowledge is one of the main purposes of ontologies. Before the creation of the SQO ontology, an exploration of already available ontologies was conducted to assess their reusability. From this exploration, the existing Analysis Data (ADA) and Core Concept Data (CCD) ontologies emerged as vocabularies which contain complementary elements usable for SQO. The ADA vocabulary is a design pattern which can be used for describing analytical data sets on a level that allows to make inferences about the kind of analysis applicable to the dataset. Specifically, the vocabulary about the structure of data has been reused. The CCD vocabulary describes geodata types representing spatial core concepts

(Kuhn, 2012) as well as further semantic distinctions relevant for geo-computation and analysis. Together, these two ontologies cover a consider a significant part of the SQO ontology model and provide a basis from which SQO can be extended. The ontology Development chapter will explain the use of the existing ontologies in more detail, as well as explaining the underlying concepts.

3.4.3 Competency Questions

Identifying the requirements of an ontology design pattern is an essential aspect in the process of designing and validating an ontology. As mentioned, one of the methods to achieve this is to develop a set of competency questions about the matter of interest (Gagnemi, 2005; Scheider & Tomko, 2016). Competency questions consist of a set of questions which are stated and replied in natural language. Competency questions support the ontology development process in two ways (Bezerra, Freitas, & Santana, 2013):

- For creating the ontology vocabulary because the questions enable the identification of the main ontology elements and the involved relationships.
- By providing a simple means to check whether the requirements of an ontology are fulfilled. This is done through the retrieval of knowledge or through inference on its axioms and answer checking.

For the purpose of developing SQO, five competency questions have been posed which will determine the scope of the ontology. The set of competency questions will be systematically answered to help identify which information is necessary to answer spatial questions. The results of the competency questions will be used to test the ontology and further develop it. Next, the competency questions are introduced.

CQ 1: Which general types of human geographic spatial questions exist?

In order to distinguish what type of workflow or query to apply when answering a spatial question, it is important for analysts to determine what question type they are dealing with. When more is known about question types and how they relate to certain queries, dataset types or spatial concepts, this would be valuable information.

CQ 2: What is the variety of spatial and temporal extents in human geographic spatial questions?

Determining the limitations of a spatial question is critical for analysts when determining the scope of their research. Knowing the boundaries of both the spatial and the temporal extent of a question provides a lot of information about how to answer the question.

CQ 3: What is the variety of spatial concepts related to spatial questions?

A distinctive property of spatial questions is that they can be formulated using spatial concepts (Kuhn, 2012). By determining which spatial concepts are related to a certain question (type), analysts can get a better understanding of how to answer that question. Therefore, it is important to investigate the variety of spatial questions.

CQ 4: What does the query of a spatial question look like?

Spatial questions are ultimately answered by a query which runs over data items of a spatial dataset. These data items consist of the goal attribute with its corresponding support. Examining the query, a formal and executable representation of a question, provides insight in the workflow as well as in the relationships between the goal attribute and support for that specific question.

CQ 5: What types of data is asked for?

For analysts, it is important to know what type of spatial dataset they are dealing with to properly answer their spatial question. Well-known spatial dataset types such as raster, vector or tessellation, in combination with information about the spatial concepts they represent, determine the type of geo-analytical operations which are applicable to the dataset.

3.5 Phase 3: Ontology Validation

The Ontology Validation phase has the primary purpose of validating SQO. The most important aspect of this phase is to create an RDF dataset containing the collection of spatial questions from scientific literature created in the Exploration phase. Then, based on the input provided by the triplifying of the data and assessment of the competency questions, SQO has been revised or extended accordingly. This is also underlined by Noy & McGuinness (2001), who state that ontology development is necessarily an iterative process. This means that new SQO classes were created or revised whenever this was required to effectively describe the information contained within the spatial questions.

3.6 Phase 4: Analysis

The Analysis phase has the purpose of analysing the RDF dataset containing the formalized spatial questions. This section will outline the methodology which was used in this process.

3.6.1 Python scripts

To investigate the RDF dataset developed in the Ontology Validation phase, several Python scripts have been developed using PyScripter. Each script has the purpose of examining specific statistics of the dataset, corresponding to the competency questions introduced in section 3.4.3. The results which are provided by these scripts have been visualized in the form of graphs or figures and will be further reflected upon in the Analysis chapter. These will provide insight in the semantic structure of spatial questions.

For the purpose of creating these scripts, the RDFLib library was used, which is a Python library for working with RDF. This library contains an RDF/XML parser which conforms to the RDF/XML Syntax Specification. The Python script which was used to analyze the variety of spatial question types can be found in listing 3.2. Several functions are defined in this script: `load_rdf`, `n_triples`, `count_elements and main`. `Load_rdf` can be used to load an RDF file. `Load_rdf` makes use of `n_triples`, which prints the number of triples in a graph. `Count_elements` is a function which is used to count the number of elements in a list. These three functions are applied in `main`, where the collection of spatial questions is loaded, a list is created which contains all the question types and the elements of this list are counted and printed. Adaptions of this script where created to retrieve other statistics and information corresponding to the remaining competency questions.

```python
import rdflib
import rdflib.plugins.sparql as sparql
import glob
#import RDFClosure
from rdflib.namespace import RDFS, RDF, OWL
from rdflib import URIRef, BNode, Literal
from rdflib import Namespace
#from urlparse import urlparse
import os

ADA = rdflib.Namespace("http://geographicknowledge.de/vocab/AnalysisData.rdf#")
SQO = rdflib.Namespace("http://geographicknowledge.de/vocab/SpatialQuestionOntology.rdf#")

"""Loads RDF data"""
def load_rdf( g, rdffile, format='turtle' ):
    #print("load_ontologies")
        #print("   Load RDF file: "+fn)
    g.parse( rdffile, format = format )
    n_triples(g)
    return g

""" Prints the number of triples in graph g """
def n_triples( g, n=None ):
    if n is None:
        print( '   Triples: '+str(len(g)) )
    else:
        print( '   Triples: +'+str(len(g)-n) )
    return len(g)

""" Counts the number of elements in a list """
def count_elements(lst):
    elements = {}
    for elem in lst:
            if elem in elements.keys():
                elements[elem] += 1
            else:
                elements[elem] = 1
    return elements

"""Loads rdf data and counts question types"""
def main():
    print ('Load questions!')
    output = rdflib.Graph()
    questions = load_rdf(rdflib.Graph(),'CombinedQuestionsFinal.ttl')
    l = []
    for q in questions.subjects(SQO.hasSpatialExtent,None):
        print (questions.value(q,RDFS.label))
        for c in questions.objects(q, RDF.type):
            print (c)
            l.append(c)
    print(count_elements(l))

if __name__ == '__main__':
    main()
```

Listing 3.2: Python script for analysing the variety of spatial question types in human geography

4. Ontology Development and Validation

This chapter presents the most important results of this research: the results of the ontology development and validation process for SQO. First, a model for the ontology is described, which briefly introduces its classes and relations. A more in-depth explanation of the vocabulary is provided afterwards. As noted by Noy & McGuinness (2001), the outcomes of the ontology development process were subject to its iterative nature. This means there was an interactive process between the creation of the ontology model and the triplification of the spatial question collection. Examples of how the collection of spatial questions was triplified into an RDF format will give further insight in this process. Both the OWL file containing SQO and the RDF dataset containing the collection of spatial questions have been published online in the following repository on GitHub:

https://github.com/jgwieleman/SpatialQuestionOntology

4.1 Formal ontology model

The SQO has been developed with the purpose of describing all the information which is necessary to answer spatial questions. With this goal in mind, a formal ontology model has been created which depicts all classes and relations within SQO, which can be seen in figure 4.1. Besides the classes and relations within SQO itself, this model also shows how the ontology extends the existing Analysis Data (ADA) and Core Concept Data (CCD) ontologies, which were briefly introduced in section 3.4.2.

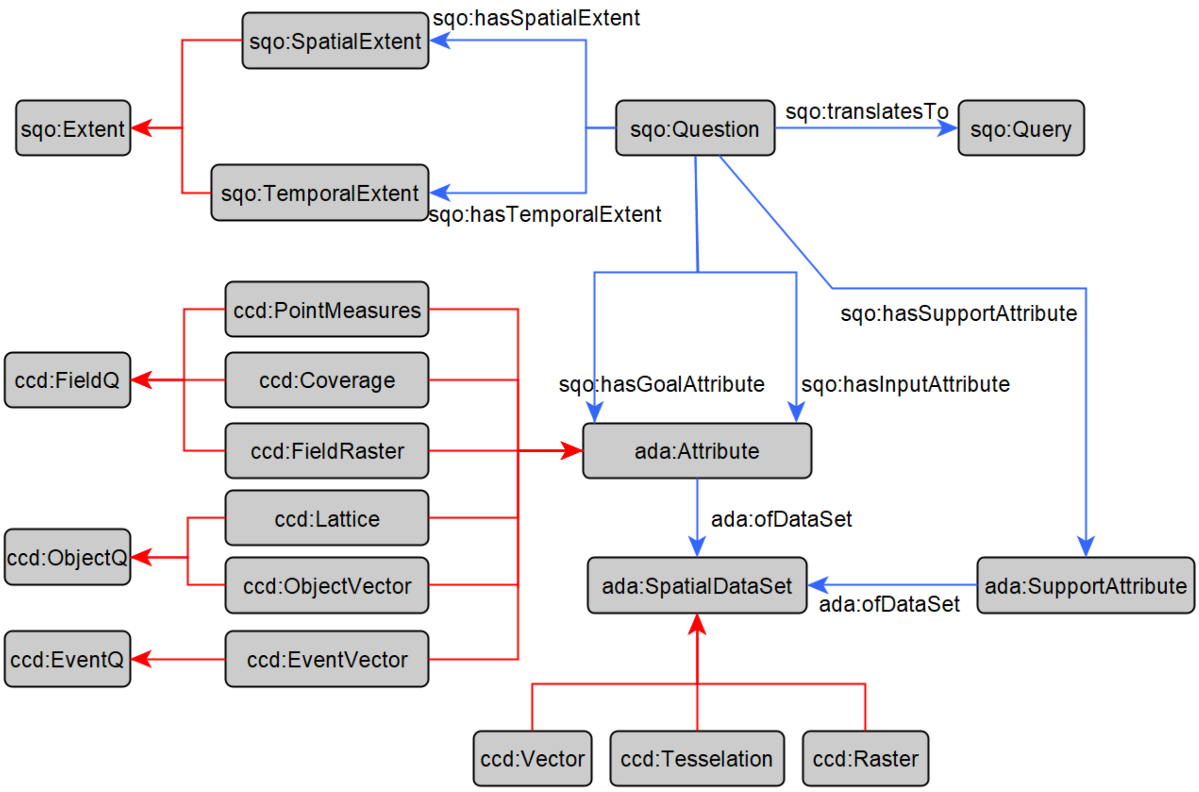

Figure 4.1: Spatial Question Ontology model. Blue arrows denote properties, red arrows denote subclasses.

The classes and properties which were reused from the ADA ontology allow the description of spatial datasets and its attributes. For interpreting the ontology model, the most important distinction to make is that a spatial data set consists of data items, which is something that identifies and combines a single observation to potentially many references. This is similar to a record in a table. Each data item can possibly have various attributes which link it to information (Scheider & Tomko, 2016). How this relates to the SQO and CCD ontologies will be further explained in section 4.2.

The classes from the CCD ontology which were reused in SQO are related to data types and the spatial concepts they represent. As lined out in chapter 2.1, core concepts of spatial information are important for the formulation of spatial questions and answers, as well as for assessing the suitability of geo-analytical tools. However, as stated by Scheider, Lemmens & Lamprecht (2019), it is of importance to note that spatial concepts are a product of human perception and measurement and should not be mistaken for the data types they represent. Scheider, Lemmens & Lamprecht (2019) have introduced a set of abstract data types which can be seen in the matrix in figure 4.2. These abstract data types essentially link the data types (raster, vector and tessellation) to the core concepts of spatial information they represent, and thus allow an assessment of which geo-analytical operations can be applied to the data. Therefore, SQO is an extension to the following CCD classes: `PointMeasure`, `FieldRaster`, `Coverage`, `Lattice`, `ObjectVector` and `EventVector`. Also, the corresponding core concept classes from CCD, object, field and event quality, have been included. The relevance of including abstract data types in SQO lies in the fact that it enables adding the type signatures to geo-analytical (GIS) operations. This allows the creation of meaningful applications and also to automate the construction of workflows (Scheider, Lemmens & Lamprecht, 2019).

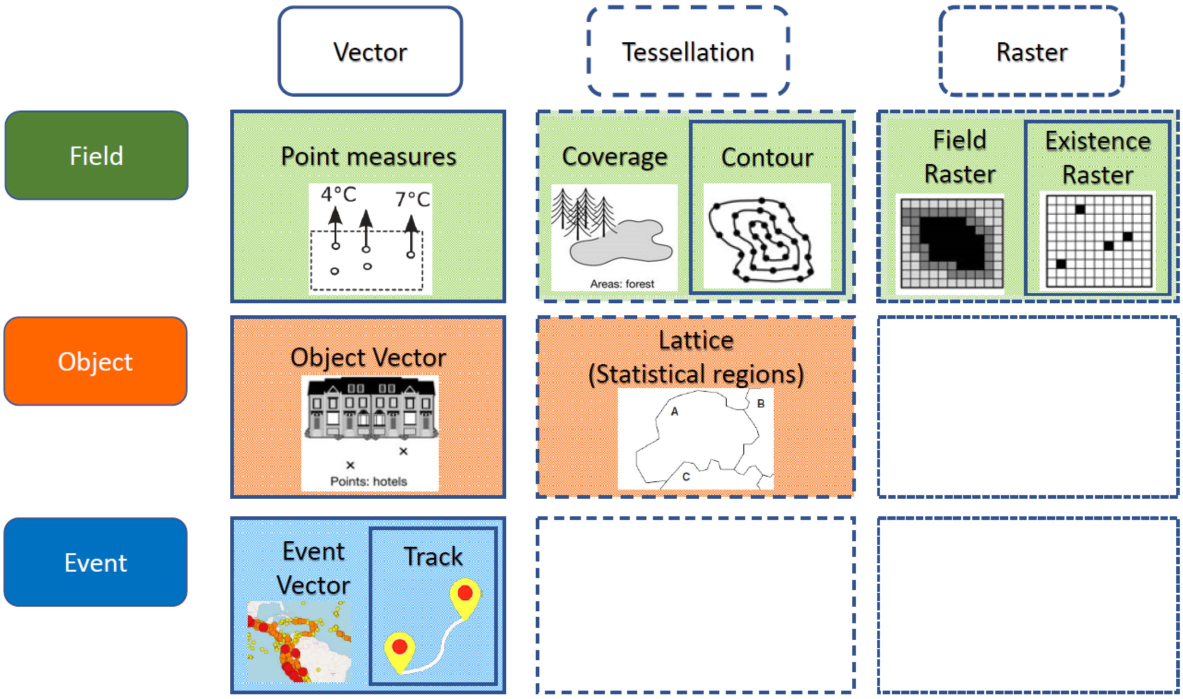

Figure 4.2: Matrix of abstract data types based on a combination of spatial concepts and geodata types (Scheider, Lemmens & Lamprecht, 2019).

4.2 Ontology Design

For the creation of SQO, several classes and properties have been developed which were briefly introduced in the ontology model in figure 4.1. This section will further elaborate upon these classes, their relations and other notable characteristics. Also, some of the underlying concepts regarding this ontology are further examined. The OWL file which contains SQO can be found in appendix B.

4.2.1 SQO Classes

SQO introduces several own classes into its vocabulary, which can be seen in table 4.1. The central class in this ontology is `Question`, which refers to a spatial question. The way to make use of this ontology is to identify a spatial question occurring in scientific literature. Then, the information which is contained within the question can be described with the other related classes.

Table 4.1: Classes of the SQO vocabulary

Class	Explanation
`Question`	Class of spatial questions
`Query`	The query to which the question translates
`Extent`	The extent of a spatial question
`SpatialExtent`	The spatial extent of a question
`TemporalExtent`	The temporal extent of a question

It can be argued that each spatial question is limited in both time and space. Every spatial question has a spatiotemporal extent for which it is attempting to find an answer. For example, the spatial question "*What is the population density per neighbourhood in the city of Utrecht in 2018?*" is limited both spatially and temporally. It is trying to find an answer for the area of Utrecht and the period of 2018. This is expressed in the `Extent` class. This class has the two important subclasses `SpatialExtent` and `TemporalExtent`. The spatial extent of a question can be seen as the area of study which the question is examining. The temporal extent can be described as the period of study which the question is examining. Together, these two subclasses define the spatiotemporal boundaries of a geo-analytical question.

The `Question` class has several subclasses, which refer to different question types. Due to the number of subclasses, these have been omitted from the ontology model above for clarity purposes. The question types were proposed based on the spatial questions retrieved from scientific literature and examples include: `Suitability, Accessibility, Cluster` and `Track`. The question type classification was created based on both the thematic and methodical aspects of the questions. Some question types are more methodical, and some are more thematic. However, each of these classes refer to a workflow/query which is unique for that specific type of question when deducing the answer based on a related spatial dataset. The different question types will be examined in more detail in chapter 5.

Queries are formal representations of questions and are executable. Queries which represent a spatial question are individuals of the `Query` class. One of the possibilities of representing these natural language questions as a query is by using SPARQL. The information about a spatial question which is described by the other classes of SQO enable the construction of a query. Of particular importance is information about the goal and support attribute and the type of question, since these determine the workflow. This will be described in more detail in section 4.3.

If a spatial question is answerable by using a GIS, this is done by analysing a spatial dataset. Well-known spatial dataset types are vector and raster datasets (Scheider & Tomko, 2016). However, another geodata type can also be distinguished: tessellations. Tessellations are a tiling of the plane into regions which jointly cover the plane and are mutually non-overlapping (Scheider, Lemmens & Lamprecht, 2019). All three geodata types are included in SQO with the class `ada:SpatialDataSet` and its subclasses.

In order to properly interpret the relations which interact with the spatial dataset class in SQO, it is of importance to examine the structure of these datasets. As noted by Scheider & Tomko (2016), a spatial dataset consists of data items. Each data item is something that identifies and combines a single observation with potentially many references. A data item can be compared to a record of database table. Such a record can possibly have various different attributes and is frequently identified by a primary key value. Scheider & Tomko (2016) argue that data items which are part of a spatial dataset are always required to have at least two types of attributes, which are a support and a measure. It is important to make a conceptual distinction between these two different roles an attribute can fulfil in a spatial dataset. Measures are observed references of a data item and a support attribute identifies the context of the observation which can be used to compare measures (Scheider & Tomko, 2016). Further explanation of these roles is provided in figure 4.3. In this example, statistical regions are the support attribute while population density is the measure attribute. This means that the population density was observed for a certain statistical region. The data is made up of various data items of which the regions can be used to compare the population densities.

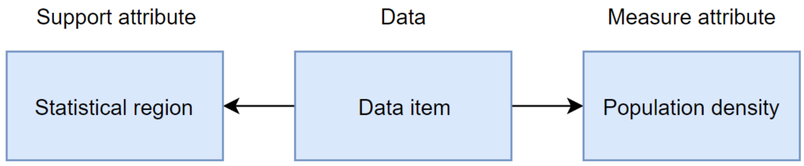

Figure 4.3: Spatial data structure which includes the measure and support attribute distinction

4.2.2 SQO properties

For the SQO ontology 6 properties have been introduced, which can be seen in table 4.2. When considering a property, the domain asserts that the subject of the property statement must belong to the specified class and the range asserts that the values of this property must belong to the specified class (Ding & Peng, 2004). The domain and range can therefore be seen as the "from" and "to" classes of a property (Scheider & Tomko, 2016). The fact that `Question` is the central class in this ontology is also reflected by the fact that each property has this class as its domain. This means that all SQO properties can only be have individuals from the `Question` class as a subject and relate to individuals of other classes as values.

Table 4.2: Properties of the SQO vocabulary

Property	Domain	Range	Explanation
`hasGoalAttribute`	`Question`	`ada:Attribute`	The goal attribute of a question
`hasInputAttribute`	`Question`	`ada:Attribute`	The input attribute of a question
`hasSpatialExtent`	`Question`	`SpatialExtent`	The spatial extent of a question
`hasTemporalExtent`	`Question`	`TemporalExtent`	The temporal extent of a question
`hasSupportAttribute`	`Question`	`ada:SupportAttribute`	The support of a question
`translatesTo`	`Question`	`Query`	The query to which the question translates

When considering spatial questions answerable by a GIS, there are always specific attributes of a spatial dataset which can used to determine the answer to the question. This is always a combination of a measure and a support attribute. For instance, when answering the question "*What is the population density per neighbourhood in the city of Utrecht in 2018*?", the population density is the measure attribute and the neighbourhoods (statistical regions) are the support attribute. The measure attribute, which can be considered to be the goal of the question, is captured in SQO by the `hasGoalAttribute` property. However, in order to answer the spatial question (by analysing the goal attribute), the support attribute is necessary for being able to compare the goal attribute's observed values. The relationship between the answer of a question and the support attribute is captured in SQO by the `hasSupportAttribute` property. These relationships between the goal and corresponding support attribute of a spatial question, their input attributes and the query resulting in the answer to a spatial question is visualized in figure 4.4.

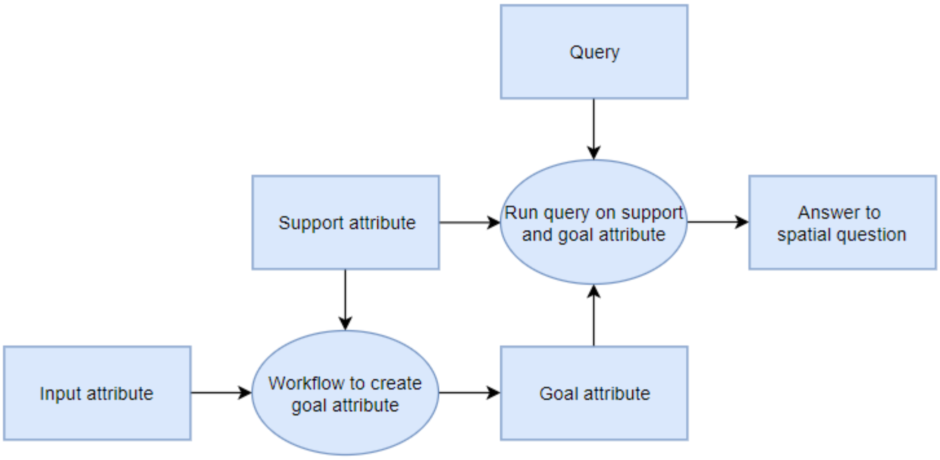

Figure 4.4: Relationships between goal, support and input attributes and the corresponding query and answer of a spatial question.

When an analyst is answering a spatial question, the goal attribute of that question is unknown and must be collected or calculated in that case. For the spatial question "*What is the population density per neighbourhood in the city of Utrecht in 2018*?", the goal attribute is the population density, as stated before. In the case this attribute is unknown it could be calculated with the population count attributes and the area attributes of the neighbourhoods (the support). The attributes of datasets which are used to calculate the goal attribute of a question can be referred to with the `hasInputAttribute` property. This relationship is also shown in figure 4.4.

An important characteristic of an input attribute is that it does not have to be a part of the same dataset as the goal attribute. This means that it can also have a divergent dataset type or abstract data type attribute. For instance, input attributes of a vector dataset can be used to calculate the goal attribute in a raster dataset. It is also noted that input attributes are per definition not directly involved in the query which results in the answering of the spatial question. However, input attributes do provide important information about a spatial question, namely about how the dataset which contains the goal attribute was or can be created. This information is important because it provides analysts with the knowledge about which geo-analytical operations they need to perform to get the goal attribute of a spatial question.

The spatiotemporal boundaries of a question can be described with the `hasSpatialExtent` and `hasTemporalExtent` properties. By describing these relations, the area of study and period of study of a spatial question can be defined. These are relationships between the central `Question` class and the important subclasses `SpatialExtent` and `TemporalExtent` of the `Extent` class.

4.3 Triplifying spatial questions

Using SQO, the collection of spatial questions from scientific literature, along with all other relevant information contained within the articles, have been triplified into an RDF dataset. Listing 4.1 contains a snippet from the RDF dataset showing a formalized question. The example spatial question "*Where are the sites located which are most suitable for the construction of marinas in Istanbul's Marmara shoreline?*" is from the article by Gumusay, Koseoglu & Bakirman (2016). In order to further explain the process of formalizing these natural language questions in RDF, more details about this example are provided below.

```
537 #Question 22: Gumusay, M.U., Koseoglu, G. & Bakirman, T., 2016
538
539 :Question22 a sqo:Suitability; #This is an suitability question
540     rdfs:label "Where are the sites located which are most suitable for the construction of marinas in Istanbul’s Marmara Sea shoreline?";
541     rdfs:comment "for a marina (harbour) site at Istanbul’s Marmara Sea shoreline";
542     sqo:hasSpatialExtent dbp:Istanbul_Province;
543     sqo:hasTemporalExtent "2018"^^time:generalYear;
544     sqo:hasGoalAttribute sqo:SuitabilityScore; #the goal attribute is a suitability score
545     sqo:hasSupportAttribute _:Grid; #This defines the support (for generating the attribute) to be a grid
546     sqo:hasInputAttribute _:Erosion;#Here, all the input attributes are listed which are used to calculate the goal attribute
547     sqo:hasInputAttribute _:LandslideRisk;
548     sqo:hasInputAttribute _:GeologicalUnsuitableArea;
549     sqo:hasInputAttribute _:SeabedSlope;
550     sqo:hasInputAttribute _:Landuse;
551     sqo:hasInputAttribute _:SeabedTansmissionLines;
552     sqo:hasInputAttribute _:SeaTraffic;
553     sqo:hasInputAttribute _:DemographicData;
554     sqo:hasInputAttribute _:DistanceToMainRoads.
555
556 sqo:SuitabilityScore a ccd:Lattice;
557     ada:ofDataset ccd:Tesselation.
558
559 _:Erosion rdfs:label "Erosion".
560 _:LandslideRisk rdfs:label "Landslide Risk".
561 _:GeologicalUnsuitableArea rdfs:label "Geological Unsuitable Area".
562 _:SeabedSlope rdfs:label "Seabed Slope".
563 _:Landuse rdfs:label "Landuse".
564 _:SeabedTansmissionLines rdfs:label "Seabed Transmission Lines".
565 _:SeaTraffic rdfs:label "Sea Traffic".
566 _:DemographicData rdfs:label "Demographic Data".
567 _:DistanceToMainRoads rdfs:label "Distance To Main Roads".
568
569 dbp:Istanbul_Province a sf:Polygon; #This is a further specification of the spatial extent of the question. The study area lies within a polygon
570     geo:asWKT "POLYGON((-77.089005 38.913574,-77.029953 38.913574,-77.029953 38.886321,-77.089005 38.886321,-77.089005 38.913574))"^^geo:wktLiteral.
```

Listing 4.1: Example of a formalized question from the RDF dataset

On line 538 in listing 4.1, it is stated that this question (22) is a suitability analysis type question. An `rdfs:label` can be used to provide a human-readable version of a resource's name. The label corresponding to this question is then added on line 539 in order to add the natural language question to the dataset as well.

The spatial extent of this question is described on lines 541 and 569-570. The area of study of this question lies within the Istanbul province in Turkey. Therefore, the spatial extent is first described with the URI `dbp:Istanbul_Province`. The URI refers to the Istanbul Province DBpedia page. DBpedia is a crowd-sourced knowledge base containing structured information from several Wikimedia projects. For the current research, the choice was made to make use of the structured information on DBpedia to describe the spatial extent of the questions since most of the spatial extents already exist within the Linked Data published on DBpedia. The resources on DBpedia are published with the prefix `http://dbpedia.org/resource/`, which is shortened to `dbp:` in the snippet. Lines 569-570 offer further specification of the spatial extent of this question. By stating that the study area lies within a polygon with corresponding coordinates the study area of the question can be described in more detail. It must be noted that the coordinates provided in the example above are not the actual coordinates of the polygon representing the Istanbul Province, since the list of actual coordinates would be too extensive to fit the purpose of this example. They merely serve as a demonstration of how SQO can be used to define the spatial extent of a question. It must be noted that a spatial question can have multiple areas of study which together make up the spatial extent. Listing 4.2 provides an example of such a question. In this case, the spatial extent consists of four separate areas.

```
#Article 3: Romanillos, G., García-Palomares, J.C. , 2018

:Question3 a sqo:Accessibility;
    rdfs:label "How to evaluate existing accessibility to public facilities in relation to the amount of resources deployed and their location?";
    rdfs:comment "For public and private primary schools in four European cities";
    sqo:hasSpatialExtent dbp:Barcelona;
    sqo:hasSpatialExtent dbp:London;
    sqo:hasSpatialExtent dbp:Madrid;
    sqo:hasSpatialExtent dbp:Rotterdam;
```

Listing 4.2: Spatial extent of a question with multiple study areas

The temporal extent of the question in listing 4.1 is the year 2018, the moment in time for which the question is trying to find an answer. The temporal extent is specified on line 542 by making use of the OWL-Time ontology. OWL-time is a useful ontology for describing the temporal extent of a question due to the fact that it allows various ways of describing the temporal properties of resources. The OWL-time vocabulary can be used to express facts about the topological relations of instants and intervals, combined with information of their durations. Also, date-time information can be expressed by using several temporal reference systems such as the conventional Gregorian calendar (Hobbs & Pan, 2017). This enables very specific description of the temporal extent of a spatial question. Something to consider is that the level of specificity offered by OWL-time greatly exceeds the specificity of the example. This is due to the fact that the article by Gumusay, Koseoglu & Bakirman (2016) does not provide enough information about the study to be more exact.

Line 543 of the example in listing 4.1 states that the goal attribute of the spatial question is a suitability score. This is basically an index assigned to each grid cell in the analysis based upon a number of different input attributes, typical for site suitability type questions. Furthermore, on line 555 it is expressed that the goal attribute is a lattice attribute and on the line after that, that it is part of dataset which is a tessellation.

Lines 545-553 describe all the different input attributes which are used to generate the goal attribute. Labels are assigned to these input attributes in lines 559-567 to provide human readable versions for when the data is analysed. As mentioned before in section 3.2.2, it must be stressed that these input attributes are not part of the same dataset containing the goal attribute and the support. They were used to calculate the goal attribute but belong to different dataset(s) (types). The dataset containing these input attributes were not included as it was not possible to extract all the necessary information about these input attributes' datasets and abstract data types from the article.

As mentioned, a spatial question can be translated to a query. This was also done for the question from the article of Gumusay, Koseoglu & Bakirman (2016) resulting in the SPARQL query in listing 4.3. In this case, the query can be written in natural language as "Which support values in the goal dataset have a goal attribute value above 0.8?". The query is a generalized representation of the question and it can be used to get the answer to the spatial question: "*Where are the sites located which are most suitable for the construction of marinas in Istanbul's Marmara shoreline?".*

```
Select ?supportvalue where {
    :Question22 sqo:hasGoalAttribute ?attribute.
    ?attribute ada:hasValue ?attributevalue .
    :Question22 sqo:hasSupportAttribute ?support.
    ?support ada:hasValue ?supportvalue .
    ?dataitem ada:hasAttribute ?supportvalue;
        ada:hasAttribute ?attributevalue.
    :Question22 sqo:hasGoalDataSet ?goaldataset.
    ?goaldataset ada:hasElement ?dataitem.
FILTER(?attributevalue > 0.8)
}
```

Listing 4.3: SPARQL query representing a spatial question

In the SPARQL query in listing 4.3, the variable `supportvalue` represents the grid cell of the goal dataset. Each grid cell is the support of a data item in this dataset and each data item also has an attribute value (the suitability score). Ultimately, to find out where these most suitable locations are, it must be checked which cells of the grid have a suitability score of above 0.8 (the threshold selected by the researchers). This is done in the query by using a filter expression to select all support values which are above this threshold.

In order for SPARQL queries to be added to the RDF dataset, they must first be converted to a suitable syntax. For this purpose, the SPIN SPARQL Syntax was used. This syntax is an RDF representation of SPARQL queries which allows the storing of the queries together with the domain model. Because all resources in the SPARQL query are represented as a valid RDF resource reference instead of a string, SPIN goes further than just providing a textual format. Another advantage of an RDF representation of SPARQL is that it aids the maintenance of hybrid models which contain both SPARQL expressions and RDF and OWL definitions (Knublauch, 2011). Ultimately, the main objective of using the SPIN syntax is to allow software tools to process SPIN RDF data into SPARQL query strings. This way, the queries can be processed for other means. Tools for editing work the other way around, turning SPARQL queries in a RDF representation (Knublauch, 2011). For the purpose of the current research, it provides the means to represent individuals of the `Query` class in a machine-readable notation.

The SPARQL query in listing 4.3 has been successfully represented in the RDF dataset using the SPIN SPARQL syntax. Listing 4.4 demonstrates how this individual of the `Query` class has been triplified. The triple representation contains all the same elements as the regular SPARQL query. This triple shows how the `translatesTo` property is used on line 574, with the query itself being stored as a blank node. The first part, lines 575-584, where the select query is initiated, also includes the filter expression and states the result variable. The various triple patterns which make up the rest of the query are stated in lines 585-616.

```
:Question22 sqo:translatesTo
[ a sp:select ;
  rdfs:comment "Which support values in the goal dataset have goal attribute value above 0.8";
  sp:Filter
          ( sp:expression
              [ a sp:ge ;
                sp:arg1 [ sp:varName "attributevalue"^^xsd:string ] ;
                sp:arg2 0.8
              ] ) ;
   sp:resultVariables
    ([ sp:varName "supportvalue"^^<xsd:string> ]);
   sp:where ([sp:object :question22;
              sp:predicate sqo:hasGoalAttribute ;
              sp:subject [ sp:varName "attribute"^^xsd:string ]
             ]
             [sp:object [ sp:varName "attribute"^^xsd:string ];
              sp:predicate sqo:hasValue;
              sp:subject [ sp:varName "attributevalue"^^xsd:string ]
             ]
             [sp:object :Question22;
              sp:predicate sqo:hasSupportAttribute;
              sp:subject [ sp:varName "support"^^xsd:string ]
             ]
             [sp:object [ sp:varName "support"^^xsd:string ];
              sp:predicate ada:hasValue;
              sp:subject [ sp:varName "supportvalue"^^xsd:string ]
             ]
             [sp:object [ sp:varName "dataitem"^^xsd:string ];
              sp:predicate ada:hasAttribute;
              sp:subject [ sp:varName "supportvalue"^^xsd:string ]
             ]
             [sp:object [ sp:varName "dataitem"^^xsd:string ];
              sp:predicate ada:hasAttribute;
              sp:subject [ sp:varName "attributevalue"^^xsd:string ]
             ]
             [sp:object :Question22 ;
              sp:predicate sqo:hasGoalDataset;
              sp:subject [ sp:varName "goaldataset"^^xsd:string ]
             ]
             [sp:object [ sp:varName "goaldataset"^^xsd:string ];
              sp:predicate ada:hasElement;
              sp:subject [ sp:varName "dataitem"^^xsd:string ]
             ] )
     ].
```

Listing 4.4: Query representing a spatial question in SPIN syntax.

5. Analysis

The development of SQO has enabled the description of information contained within spatial questions. This has led to the creation of an RDF dataset containing a collection of questions from human geographic literature. The aim of this chapter is to analyse the spatial questions in this dataset by systematically answering the competency questions which were created for this ontology. These results signify how the capabilities of SQO can be applied.

5.1 Types of spatial questions

A taxonomy of question types has been created in SQO for spatial questions based upon the method and thematic aspects related to the goal (attribute) of the questions. These question types are subclasses of the `Question` class in SQO and can be found in table 5.1. As explained in section 4.2.1, these question types were determined during the creation of the RDF dataset. These are based upon the thematic and methodical characteristics of the questions which were included in the dataset.

Table 5.1: Spatial question types and explanation

Spatial question type	Explanation
`Accessibility`	Questions which analyse spatial accessibility
`Accuracy`	Questions which analyse the accuracy of a spatial phenomenon
`Granularity`	Questions which analyse granularity
`LandUse`	Questions which analyse a spatial phenomenon based on land use (change)
`Suitability`	Questions which analyse spatial suitability
`Track`	Questions which analyse event tracks
`NetworkAnalysis`	Questions which analyse networks
`Regression`	Questions which use regression to analyse a spatial phenomenon
`SpatiotemporalTrend`	Questions which analyse a spatiotemporal trend
`SpatialDistribution`	Questions which analyse the distribution of a spatial phenomenon
`Cluster`	Questions which analyse clusters (subclass of `SpatialDistribution`)
`PopulationDistribution`	Questions which analyse population distributions (subclass of `SpatialDistribution`)

As mentioned, the main determinant for the classification of a question was its goal. Most of the time, this related directly to its goal attribute. An example of this is the question *"How can an urban spatial design model be applied for modeling the settlement expansion in Addis Ababa?"* from the article by Abo-El-Wafa, Yeshitela & Pauleit (2018). The goal attribute of this question is land use and therefore, it belongs to the land use question type. However, it was not always possible to directly categorize a question based upon its goal attribute, as will be explained in the next paragraph.

It must be noted that the question type classes are not mutually exclusive, meaning that a spatial question can share the characteristics of multiple question types. This is often the case for accuracy and granularity questions. One example for this is the spatial question: '*How can a grid size suitability evaluation method be developed for grid-based population data?'* from the article of Dong, Yang, Cai & Huang (2017). This question was classified as a granularity question, since investigating the grid size suitability is the primary goal of the question. However, it also has the characteristics of a population distribution question since its goal attribute is related to population distribution. Although in this case, the granularity aspect of the goal attribute is the true aim of the question.

The spatial question type of all questions in the dataset were determined based upon the goal of the question. The variety of spatial question types can be seen in figure 5.1. This chart provides an overview of all the occurrences of spatial question types in the RDF dataset.

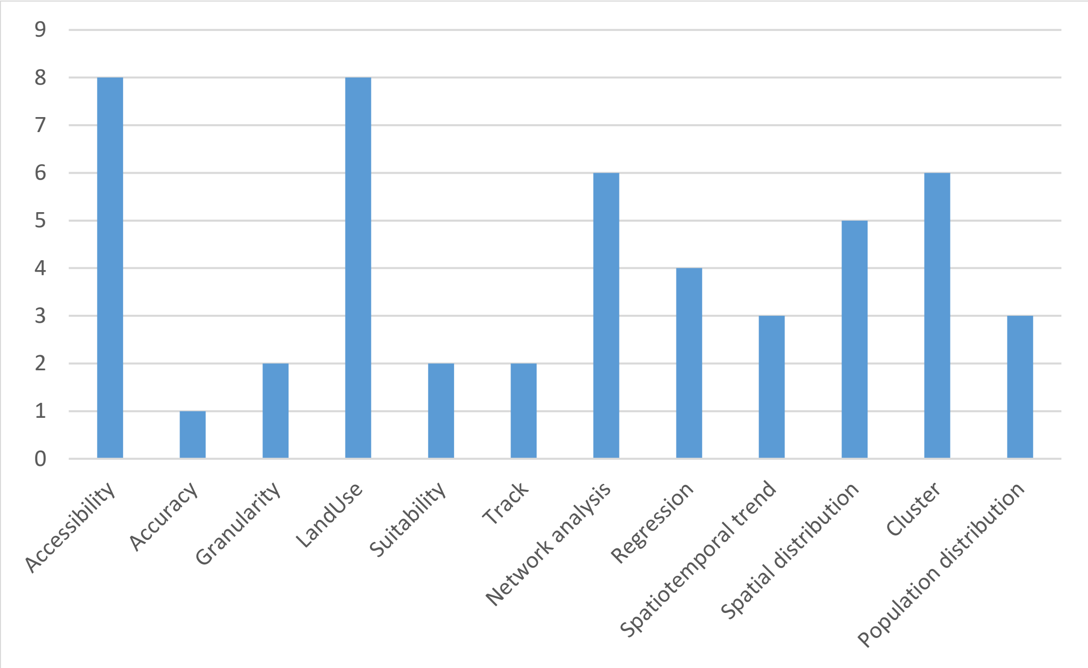

Figure 5.1:Variety of spatial question types published in human geographic scientific literature.

From these statistics, it is clear that questions which investigate the distribution of a spatial phenomenon are most common in scientific literature about human geography. Combined, the three question types spatial distribution (5), and its subclasses cluster (6) and population distribution (3) were recorded 14 times (28% of total). Other much occurring question types are accessibility and land questions, which both were present 8 times (16% of total) in the dataset. The least-occurring question type was accuracy, which was only recorded once (2% of total).

5.2 Spatial and temporal extent

SQO enables the description of the spatiotemporal limits of question through the spatial and temporal extent classes. By doing so, it allowed the definition of study areas and periods for the questions in the current dataset. The spatial and temporal extents of the questions in the RDF dataset will be explored in this section.

5.2.1 Spatial extent

For analysts, knowing the exact spatial limits for which a question is trying to find an answer is critical. SQO currently enables the description of the study area in two ways: by referring to an URI from DBpedia (could be other URI's as well) and by formalizing the coordinates of the spatial extent. In total, the 50 spatial questions have 69 spatial extents. This is due to the fact that that a single spatial question can have multiple study areas.

Possible applications of the spatial extent class are that it can be used for geo-analytical purposes when answering the question or to create a map of the study area. An example map of a spatial extent from the RDF dataset is provided in figure 5.2. It shows the study area of the question '*What is the urban population in China based on radiance corrected DMSP-OLS nighttime light and land cover data?*', from the article by Li & Zhou (2018), which is China.

Figure 5.2: Spatial extent of the spatial question from Li & Zhou (2018).

Furthermore, an overview has been created of all the spatial extents which are present in the dataset. This map can be found in figure 5.3. In this map, the 69 spatial extents are represented by graduated symbol on top of their central coordinates.

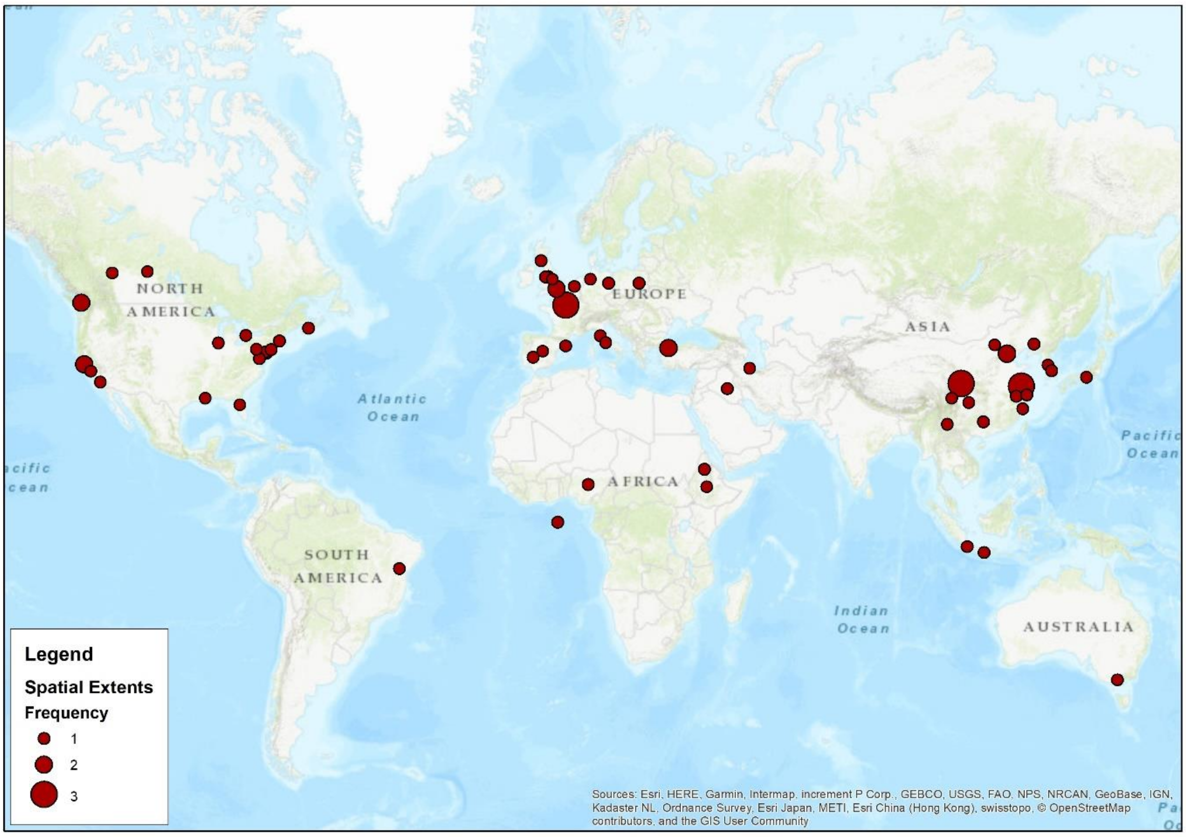

Figure 5.3: Spatial extents of spatial questions from human geography.

As can be seen in figure 5.3, the spatial extents of the spatial questions from human geography are spread across all continents except Antarctica. Also, there is a spatial extent which spans the entire earth, of which the centroid is located southwest of Africa. However, the data suggests that there are three areas which are heavily studied, these are Europe (18 spatial extents), China (19 spatial extents) and North America (19 spatial extents). When interpreting these statistics, it is important to consider the fact that several questions have multiple study areas, which slightly skew these results. For instance, the spatial question *"How to evaluate existing accessibility to public facilities in relation to the amount of resources deployed and their location?"* from the article by Romanillos & García-Palomares (2018) has four study areas: Barcelona, Rotterdam, London and Madrid.

5.2.2 Temporal extent

Besides the spatial limits, the temporal limits of a question are also of importance as these determine the temporal resolution of the question. The OWL-Time ontology was used to describe the temporal extents of the spatial questions in the RDF dataset. Even though this ontology can be used to describe quite detailed periods (e.g. months, days, hours and minutes), most temporal extents in scientific literature were described in years. Figure 5.4 provides an overview of the temporal extent of the questions of which the extent was a single year.

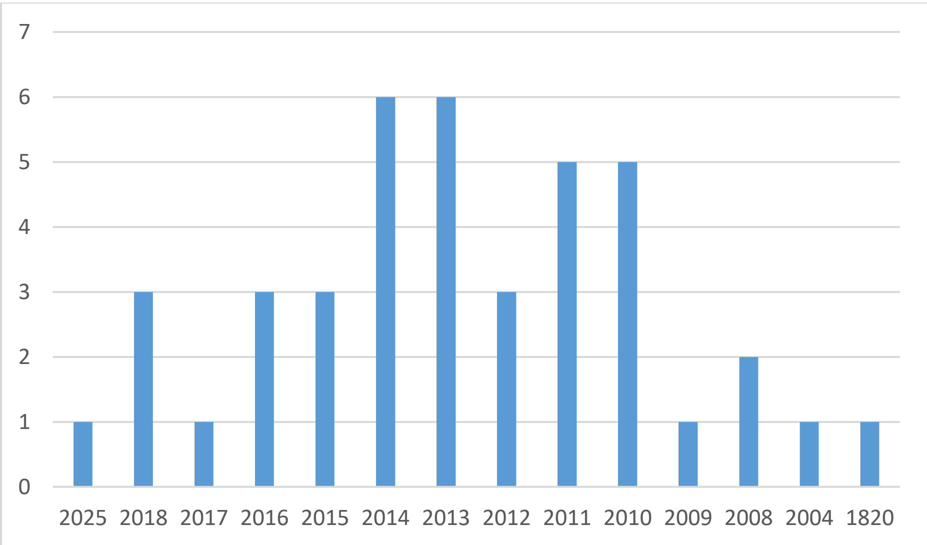

Figure 5.4: Temporal extent of spatial questions in human geography (single years).

As can be seen in figure 5.4, the temporal extents of most of the articles were spread over the years between 2018 and 2008. Within this period of time, 2014 and 2013 are the years with the most occurring temporal extents, 6 each. The years of 2017 and 2009 are the least occurring temporal extents. Noticeable temporal extents included in the RDF dataset are 2025 and 1820. The question from the 2025 temporal extent is: *"How can an urban spatial design model be applied for modeling the future settlement expansion in Addis Ababa?".* The question from the 1820 temporal extent is *"What was the historical spatial land use pattern for Jiangsu Province in the mid-Qing Dynasty?".* What can be concluded from this spread is that, beside these two outliers, most articles have a temporal extent that is several years prior to their date of publishing.

Furthermore, there were also a number of questions where the temporal extent covered a period which spanned multiple years. These temporal extents were 2012-2016, 2008-2010, 2007-2010, 2000-2010, 1993-2007, 1990-2010 and 1990-2006. All of these temporal extents only occurred once.

5.3 Query of a question

A query is a generalized representation of the question and it can be used to get the answer to the spatial question. As explained in section 4.3, SQO can be used to formalize a natural question in a SPARQL query form. Because section 4.3 elaborated on how the spatial question queries look like in SPARQL and SPIN syntax, this will only be briefly discussed in the current chapter. The example SPARQL query for the question "*Where are the sites located which are most suitable for the construction of marinas in Istanbul's Marmara shoreline?",* is shown again in listing 5.1. This query was added to the RDF dataset using the SPIN syntax such as shown in listing 4.4.

```
Select ?supportvalue where {
    :Question22 sqo:hasGoalAttribute ?attribute.
    ?attribute ada:hasValue ?attributevalue .
    :Question22 sqo:hasSupportAttribute ?support.
    ?support ada:hasValue ?supportvalue .
    ?dataitem ada:hasAttribute ?supportvalue;
        ada:hasAttribute ?attributevalue.
    :Question22 sqo:hasGoalDataSet ?goaldataset.
    ?goaldataset ada:hasElement ?dataitem.
FILTER(?attributevalue > 0.8)
}
```

Listing 5.1: SPARQL query representing a spatial question.

This example shows how spatial questions can be translated to queries. The relation between the goal and support attribute is the foundation of all queries representing spatial question. For most questions, the goal attributes of the measures will have to be compared to formulate an answer to the question. However, besides the support-based goal attribute comparison which is present in all queries, the query needs to be further adapted depending on specific characteristics of the (type of) question. In listing 5.1 for example, a filter expression (of all goal attribute values above 0.8) was added to select the most suitable sites. It can be argued that every suitability question contains a filter expression, but the height of the filter statement (0.8) is unique for each spatial question. Together, the relationships between the support attribute, goal attribute and unique question (type) characteristics determine how a spatial question can be translated to a query.

5.4 Answer information

For analysts, it is important to know what type of spatial dataset they are dealing with to properly answer their spatial question. Well-known spatial dataset types such as raster, vector or tessellation, in combination with information about the spatial concepts they represent, determine the type of geo-analytical operations which are applicable to the dataset.

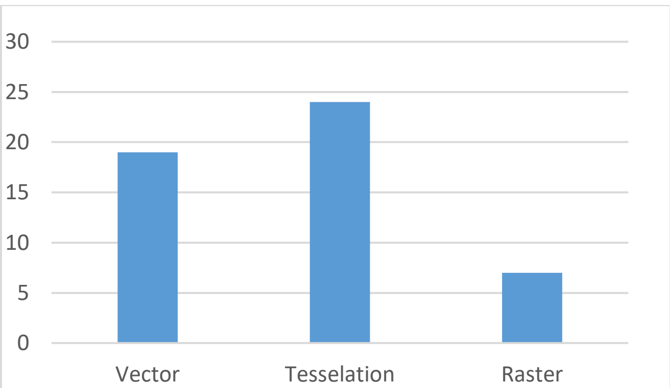

Figure 5.5: Dataset types of the goal attribute for spatial questions in human geography.

Figure 5.5 shows the incidence of different dataset types for the given spatial questions in human geography. Of the 50 goal attributes in the dataset, 19 belonged to vector, 24 to tessellation and 7 to raster datasets. These results suggest that vector and tessellation datasets are used much more in the field of human geography than raster datasets. However, these are the dataset types which contain to the goal attributes of the answers which were formulated by the authors of the articles. It must be noted that in some cases, there may be multiple ways to answer a spatial question. These results only reflect the choices made in the given answer but there may be various possible answers and corresponding dataset types and related to the same question.

5.5 Spatial concepts of a question

The core concepts of spatial information proposed by Kuhn (2012) are a vital aspect of the SQO ontology. Describing the spatial concept related to the goal attribute of a question provides much insight in the intended answer, as well as in the geo-analytical operations which are required to arrive at that intended answer. Figure 5.6 shows the occurrence of three core spatial concepts related to the goal attributes of questions in the RDF dataset.

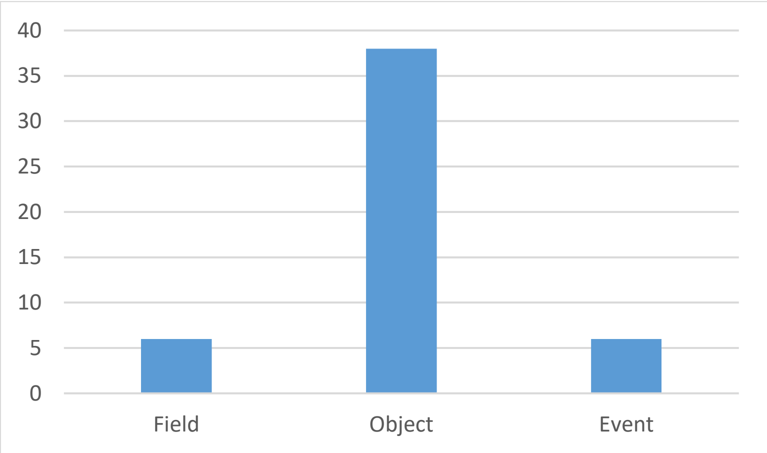

Figure 5.6: Spatial concepts of goal attributes in spatial question in human geography.

Object is by far the most-occurring spatial concept related to goal attributes in the spatial question dataset, with 38. Field and event were both recorded 6 times. However, several things must be noted when assessing these numbers. Firstly, that the three other spatial concepts, network, accuracy and granularity, were not yet taken into account in this analysis. When the spatial concept of network was related to the goal attribute of a question, it was formalized in the dataset as a special case of object due to the fact that this was not yet in the CCD ontology. Instead, the network spatial concept was captured with network analysis question type and this was also reflected in the corresponding support attributes. In total, the spatial concept of network occurred 6 times, this is included in the 38 objects. Accuracy and granularity were not included, because when the purpose of a spatial question is to investigate one of these quality concepts, it is about the accuracy or granularity of a field, object, event or network. However, there were only 2 granularity (of 1 object and 1 raster) questions and 1 accuracy (of 1 object) question.

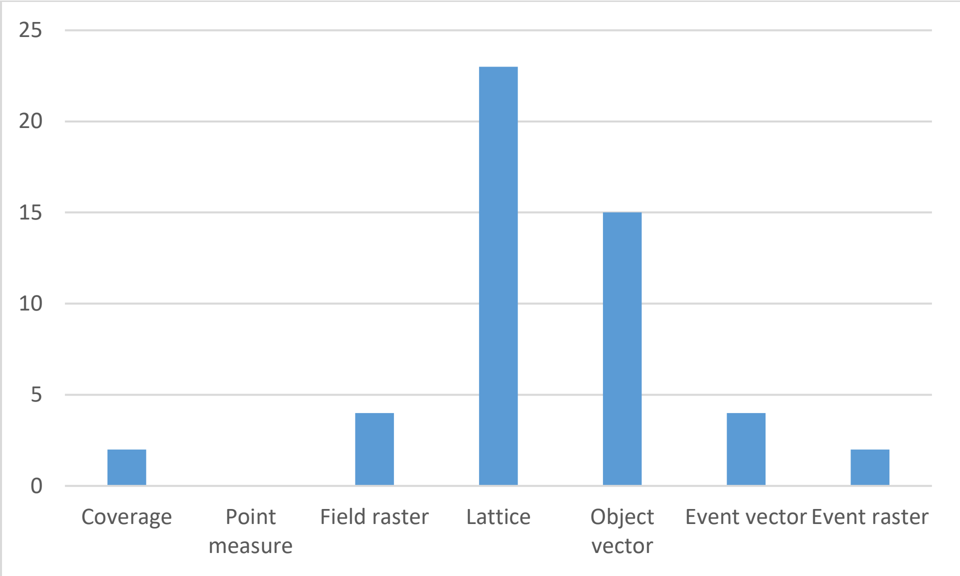

Figure 5.7: Abstract data types of goal attributes for spatial questions in human geography.

When coupled with the type of dataset, spatial concepts can be used to determine which abstract datatype a goal attribute belongs to. The abstract datatypes can be formalized using SQO and this was done for the spatial questions in the RDF dataset. Figure 5.7 shows the occurrences of abstract data types. The graph suggests that lattices (23) and object vectors (15) may be the most dominant abstract datatypes in spatial questions from human geography. Furthermore, point measure are not present at all in the dataset.

6. Conclusion

The key to understanding the semantic structure of spatial questions lies in understanding the relations between the different pieces of information which are contained within a question. Together, these pieces of information embody the meaning of a spatial question and provide insight in how to answer the question. These relations have been represented in the Spatial Question Ontology (SQO). By making use of this ontology, spatial questions from the field of human geography can be translated into a machine-readable format. Developing this ontology ultimately led to fulfilling the main goal of this research:

Empirically investigate spatial questions and the involved spatial concepts in order to develop an ontology which describes geo-analytical questions.

By answering the three sub-questions, an answer to the main research question is formulated, as well as fulfilling the goal of this research.

1. *What is the variety of spatial questions asked in Human Geography?*

As has been stated Jarvinen (2004) and Onwuegbuzie & Leech (2006), the research question is the essential factor when selecting a proper research method. Examining what the variety of spatial questions is in human geography, proved to be a valuable starting point in the research process. On the one hand, this provided the input which was necessary for improving and validating the ontology by supplying a collection of spatial questions. Furthermore, it gave helpful insight in the type of spatial questions which are currently posed in scientific literature.

Concerning the spatiotemporal boundaries of the questions investigated in this research, several conclusions can be drawn. Firstly, the data suggests that the study areas of the analysed spatial questions are quite spread out, covering all continents. However, the most densely researched areas in this collection of questions from human geography are Europe, China and North America. This might indicate that the use of GIS in human geography is more common in these areas and it would be interesting to investigate whether these clusters persist in larger scale investigations. Furthermore, regarding the temporal boundaries of the investigated spatial questions, the data suggests that most questions are related to the period the scientific articles were published (2018-2014) or the period shortly before (2013-2008). To find out whether predictive spatial questions about the future or historic spatial questions are more frequent in human geography, would be a possible future appliance of SQO together with a larger and more representative collection of spatial questions.

Based upon the analysis of the collection of spatial questions, it can be concluded that there is a diverse variety of spatial question types. However, some are more prominent in scientific literature, such as the questions which study the distribution of spatial phenomena (and its subtypes cluster analysis and population distribution). Other question types which are common include accessibility analysis, land-use analysis and network analysis questions. Questions types which occurred less frequent are: granularity, accuracy and suitability analysis questions. However, besides giving an overview of the spatial questions which are currently being asked, being able to categorize spatial questions also helps to determine the proper research methodology for that question. Ultimately, this is one of the main goals for which SQO can find future applications.

2. *Which spatial concepts are relevant for spatial questions in human geography?*

The core concepts of spatial information put forward by Kuhn (2012) and Kuhn & Ballatore (2015) proved to be a central aspect when determining the meaning and characteristics of a natural language spatial question. The base concept location, the content concepts field, object, network and event and the quality concepts granularity and accuracy are all critical when formulating spatial questions. These spatial concepts can also be used to determine whether a certain GIS tool or dataset is suitable for answering a spatial question. Therefore, they form an important element of the SQO ontology.

When examining the analysis of spatial concepts related to the goal attributes of the collected spatial questions from human geography, several observations can be made. First, that objects appear to be the most occurring spatial concept. Nevertheless, also the spatial concepts field, network and event occur quite frequently in scientific literature. The two least occurring concepts were granularity and accuracy. These outcomes suggests that a distinction can be made regarding the incidence of content concepts and quality concepts. Namely, that goal attributes related to content concepts of spatial information appear to be far more prominent in the currently investigated scientific literature.

However, it is important to note that the results from the analysis discussed above were solely based on the spatial concept related to the goal attribute of the investigated questions. It can therefore not be stated that these are the most relevant for spatial questions in human geography. In order to investigate this, it would also be necessary to examine spatial concepts related to, for example, input attributes. Furthermore, a larger collection of spatial questions would be required which is representative for human geographic literature. These things were both beyond the scope of the current research, but would be interesting opportunities for future research.

Another argument that has to be made here is that three of the spatial concepts, location, accuracy and granularity, can be considered relevant for each spatial question. However, the degree in which they have an impact on the question varies from question to question. For example, accuracy was only related to the goal of the question in two cases, but the accuracy of spatial information in general is relevant for all questions. This nuance has not directly been incorporated in SQO and the results in the analysis chapter, but is important to be aware of and it could be possible to cover this in future versions of the ontology.

3. *How can an ontology be developed which covers the semantic structure of spatial questions in human geography?*

Given the fact that there is an unlimited amount of spatial questions that can be posed, there is a need to find common structures that repeat among questions as well as answers in order to handle this variety. As a lightweight geospatial ontology, SQO makes it possible to describe these structures within a spatial question. Questions types which help determine the GIS-workflow characteristics, spatial concepts which help formulate a question and its answer and abstract data types which combine these spatial concepts to geometric dataset types in order to determine applicable geo-analytical operations. These are all common structures within spatial questions, which are captured by SQO, together with some of their possible uses.

This ontology has been developed from an answer-oriented approach, which means that it is focussed primarily on the information within a question that is necessary to formulate an answer. However, when evaluating this ontology, it is besides focussing on what can be described with SQO, also important to reflect on what is not being covered by SQO. One of the elements of spatial questions which is currently not being represented in SQO is the syntactical aspect. Being able to describe interrogative syntax with this ontology would be a useful addition, as this could open up possibilities of distinguishing question characteristics based upon the question syntax. Another element of spatial questions which is missing is the distinction between the thematic subject and the applied method when categorizing the questions. In the current form of the ontology, these aspects overlap in the question type categorization, which enabled a straightforward classification process. Possibly however, these two aspects could be separated to allow a more detailed description of spatial questions in the future.

SQO enables the description of the information contained within spatial questions into a machine-readable format which could possibly be used as a basis for automated reasoning. Enabling a machine to find out whether a resource is suitable for a specific geo-analytical task is one of the most promising possible applications of this ontology. With these possibilities, this ontology has the potential to play a central role in the development of question-based analysis.

7. Discussion

This thesis has investigated the semantic structure of spatial questions, with the geospatial ontology SQO as one of the main research outcomes. However, every research project and its outcomes are influenced by decisions made throughout the entire process. Looking back, some of the most influential decisions ought to be discussed.

Firstly, the methodology which was used to gather the collection of spatial questions from human geographic literature had a distinct impact on all the following research phases. As mentioned before, manually deducing what the spatial question within the articles was, is an arbitrary process. This may have led to the fact that some questions may not be a proper reflection of the authors intention. Even though this was the only feasible way to collect enough questions, it is still considered to be a limitation of the research.

When examining the data collection process itself, it must be noted that the results in the Analysis chapter were greatly determined by the search query used on Scopus. Even though the query reflects the most important themes within human geography, it is possible that several areas of study were not represented enough in the search. In order to be sure that these results are representative for the field of human geography, more elaborate research needs to be undertaken. This could be done by experimenting with different search queries or investigating larger question samples.

For the list the core concepts of spatial information, arguments can be made for using other concepts than the ones proposed by Kuhn (2017). Point in case, Kuhn himself opted for other lists of spatial concepts in Kuhn (2012) and Kuhn & Ballatore (2015). This shows that it is important to always be critical on which concepts can be considered 'core' to spatial information, also for future work on the matter.

This thesis regarding the semantic structure of spatial questions could prove to be the basis for further research directions. Most notably, regarding the development of applications for matching spatial questions with the right geo-analytical tools or datasets. If SQO could be applied in such applications, that would mean significant progress in this field of study.

Developing ontologies is an iterative and ongoing process, which means there is still plenty to improve upon in the SQO ontology. One possible opportunity for improvement would be to continue with the process of ontology development by validating it on new and more spatial questions. By further validating and optimizing SQO, the quality of the ontology could be improved.

In its current state, the taxonomy of spatial questions was based on spatial questions from human geography. Expanding the scope of the ontology by also incorporating spatial questions from other scientific areas in which spatial questions are present would be another research direction for SQO. Also, SQO could be transformed from a light-weight ontology into an ontology with more formal logic. By doing so, SQO would become a more extensive but also more powerful ontology which might ultimately pave the path for question-based tool interaction and data analysis.

References

Antoniou, G., & Van Harmelen, F. (2004). *A semantic web primer*. MIT press.

Berners-Lee, T., Hall, W., Hendler, J., O'Hara, K., Shadbolt, N., & Weitzner, D. (2006). A framework for web science. *Foundations and Trends in Web Science*, *1*(1), 1-130.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific american*, *284*(5), 28-37.

Bezerra, C., Freitas, F., & Santana, F. (2013). Evaluating ontologies with Competency Questions. In *Proceedings 2013 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IATW 2013*, *3*, 284–285.

Cyganiak, R., Wood, D., & Lanthaler, M. (2014). Concepts and Abstract Syntax. Retrieved December 4, 2018, from: https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/Overview.html

Ding, Z., & Peng, Y. (2004). A Probabilistic Extension to Ontology Language OWL - In *Proceedings of the 37th Hawaii International Conference On System Sciences (HICSS-37), 2004.*

Donnelly, M. B. (2005). Relative places. *Applied Ontology*, *1*(1), 55–75.

Gangemi, A. (2005). Ontology Design Patterns for Semantic Web Content. In *International semantic web conference* (pp. 262-276). Springer, Berlin, Heidelberg..

Gao, S., & Goodchild, M (2013). Asking spatial questions to identify GIS functionality. In *Fourth International Conference on Computing for Geospatial Research and Application, 2013*, 106–110.

Han, X., Feng, Z., Zhang, X., Wang, X., Rao, G., & Jiang, S. (2016). On the Statistical Analysis of Practical SPARQL Queries. In *Proceedings of the 19th International Workshop on Web and Databases*, 6, 1-6.

Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. *Synthesis lectures on the semantic web: theory and technology*, *1*(1), 1-136.

Hobbs, J., & Pan, F. (2017). Time ontology in OWL. Retrieved March 10, 2019, from: https://www.w3.org/TR/owl-time/

Hoyland, C. A., Adams, K., Tolk, A., & Xu, L. (2014). The RQ-Tech methodology: a new paradigm for conceptualizing strategic enterprise architectures. *Journal of Management Analytics*, *1*(1), 55–77.

Janelle, G. & Goodchild, M. (2004). *Spatially integrated social science*. Oxford University Press.

Jarvinen, P. (2004). Research Questions Guiding Selection of an Appropriate Research Method. In *Proceedings of European Conference on Information Systems*, 2000, 3-5.

Knublauch, H. (2011). SPIN - SPARQL Syntax. Retrieved April 4, 2019, from: https://www.w3.org/Submission/2011/SUBM-spin-sparql-20110222/

Knublauch, H., Fergerson, R. W., Noy, N. F., & Musen, M. A. (2004). The Protégé OWL plugin: An open development environment for semantic web applications. In *International Semantic Web Conference* (pp. 229-243). Springer, Berlin, Heidelberg.

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

Kuhn, W. (2017). Core Concepts of Spatial Information. Retrieved December 12, 2018, from http://spatial.ucsb.edu/core-concepts-of-spatial-information/

Kuhn, W., & Ballatore, A. (2015). Designing a language for spatial computing. In *Agile 2015* (pp. 309–326). Springer, Cham.

Maedche, A., & Staab, S. (2001). Ontology Learning for the Semantic Web. *IEEE Intelligent Systems*, *16*(2), 72–79.

McGuinness, D., & Van Harmelen, F. (2004). *OWL Web Ontology Language Overview*. 1–22.

Miltiades, L., & Angelides, D. (2016). *Spatial Planning and Semantics: An Innovative Spatial Decision Support System for Land-Use Planning*. 2012, 45–46.

Noy, N. F., Sintek, M., Decker, S., Crubézy, M., Fergerson, R. W., & Musen, M. A. (2001). Creating Semantic Web contents with Protége-2000. *Intelligent Systems, IEEE*, *16*(2), 60–71.

Noy, N., & McGuinness, D. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*.

Onwuegbuzie, A., & Leech, N. (2006). Linking Research Questions to Mixed Methods Data Analysis Procedures 1. *The Qualitative Report*, *11*(3), 474–498.

Perez, J., Arenas, M., & Gutierrez, C. (2006). Semantics and Complexity of SPARQL. *ACM Transactions on Database Systems (TODS)*, *34*(3), 1–45.

Scheider, S., Gräler, B., Pebesma, E., & Stasch, C. (2016). Modeling spatiotemporal information generation. *International Journal of Geographical Information Science*, *30*(10), 1980–2008.

Scheider, S. , Lemmens, R & Lamprecht, A. (2019). Semantic data type signatures for representing spatial core concepts in GIS operations on spatial layers. In *14th Conference on Spatial Information Theory.* Retrieved April 29, 2019, from: https://www.overleaf.com/5468495663tvdtpcxsvppy.

Scheider, S., & Tomko, M. (2016). Knowing Whether Spatio-Temporal Analysis Procedures Are Applicable to Datasets. In *FOIS* (pp. 67-80).

Tateosian, L. (2015). Beginning Python. In *Python For ArcGIS* (pp. 13–35). Springer, Cham.

Vahedi, B., Kuhn, W., & Ballatore, A. (2016). *Question-Based Spatial Computing - A Case Study*. In *Geospatial Data in a Changing World* (pp. 37-50). Springer, Cham.

Appendix A: Spatial questions collection

Nr.	Author(s)	Year	Spatial question
1	Qiao, Cao, Liu & Wu.	2018	How can the EPIC-model be applied to assess the simulation accuracies at different scales?
2	Abo-El-Wafa, Yeshitela & Pauleit	2018	How can an urban spatial design model be applied for modelling the settlement expansion in Addis Ababa?
3	Romanillos, García-Palomares	2018	How to evaluate existing accessibility to public facilities in relation to the amount of resources deployed and their location?
4	Wong, Huangfu & Hadley	2018	To what extent are residential sites for people with intellectual disabilities and physical disabilities spatially dispersed in Pennsylvania?
5	Li & Zhou	2018	What is the urban population in China based on radiance corrected DMSP-OLS night-time light and land cover data?
6	Pei, Niu, Wang, Wu & Jiang	2018	Which spatiotemporal dynamics of carbon emissions and carbon sinks exist in South China?
7	Al-Yasery, Almuhanna & Al-Jawahery	2018	Where are the sites which are most suitable to function as metro stations located in Karbala?
8	Calovi & Seghieri	2018	What is the spatial access to outpatient services relative to the demand in Tuscany?
9	Amri, Wiguna & Yunus, R.	2018	What is the population distribution in the Banten province in Indonesia?
10	Jiang, Xiong, Wang, Zhang & Ren	2018	To what extent are there regional differences in population distribution in the Yunnan province, China?
11	Susilo	2017	Which are the determinant factors in land use change in Yogyakarta?
12	Calka, Nowak Da Costa & Bielecka	2017	How can high resolution population data together with cadastral data be used in risk assessment?
13	Sousa, Pitombo, Rocha, Salgueiro & Delgado	2017	Which regions have the highest incidence of violence in public transportation?
14	Hajilo, Masoom, Motiee Langroudi, Sabokbar & Pennington-Gray	2017	How are small touristic businesses spatially distributed in the Gilan province, Iran?
15	Shah, Milosavljevic & Bath	2017	What is the geographic accessibility to primary healthcare services in relation to the distribution of seniors in Saskatchewan and Alberta?
16	Shen & Chai	2017	To what extent are there differences in spatial-temporal fixity and flexibility between individuals in Beijing?
17	Dong, Yang, Cai & Huang	2017	How can a grid size suitability evaluation method be developed for grid-based population data?
18	Barton, Weil, Jackson & Hickey	2017	What is the influence of neighbourhood violent crime on fear of crime?
19	Cai, Huang & Song	2017	How can the structure of polycentric cities be identified using multi-source geospatial big data?
20	Esch, Heldens, Hirner, Dech & Strano	2017	How can human settlements be identified using remote sensing on a global scale?
21	Pan, Zhao, Wang & Shi	2016	What is the spatial access to private and public hospitals in Sichuan, China?
22	Gumusay, Koseoglu & Bakirman	2016	Where are the sites located which are most suitable for the construction of marinas in Istanbul's Marmara Sea shoreline?
23	Bock & Sester	2016	How can parking availability maps be improved by using sensor data from nearby roads in San Francisco?
24	Saghapour, Moridpour & Thompson	2016	What is the spatial accessibility to public transport in Melbourne, Australia?
25	Paldino, Kondor, Bojic, González & Ratti	2016	What are the urban temporal patterns of individuals in big cities based upon geo-tagged photography?
26	Chen, Q., Mei, K., Dahlgren, R.A., Gong, J., Zhang, M.	2016	What is the impact of land use and population density on surface water quality in the Wen-Rui Tang River watershed in China?
27	Jin, Pan, Yang, Bai & Zhou	2016	What was the historical spatial land use pattern for Jiangsu Province in the mid-Qing Dynasty?

28	Osama & Sayed	2016	What is the impact of bike network indicators on cyclist safety using macro-level collision prediction models?
29	Chen, Zhou, Chen, Bi & Kinney	2016	How does urbanicity affect the vulnerability to heat-related mortality?
30	Sun, Xiu, Pang & Song	2016	How are economic activities clustered within the metropolitan area of Shenyang?
31	Luan, Law & Quick	2015	What are the spatiotemporal patterns of RHFA (relative healthy food access) in the region of Waterloo, Canada?
32	Buck, Kneib, Tkaczick, Konstabel & Pigeot	2015	Which factors of the built environment influence the physical activity of children?
33	Ho, Knudby & Huang	2015	How can a method be developed for mapping heat risks at different scales?
34	Aydinoglu, Senbil, Saglam & Demir	2015	How to determine the best locations of parking, depending on parking demand?
35	Mentis, Welsch, Fuso Nerini, Bazilian & Rogner	2015	How can a GIS-based approach be developed for electrification planning?
36	Sung & Lee	2015	What is the relationship between the residential built environment and walking activity?
37	Mao, Stacciarini, Smith & Wiens	2015	How to quantify individuals' rural experience using household travel surveys and geographic information systems?
38	Gutiérrez Gallego, Naranjo Gómez, Jaraíz-Cabanillas, Ruiz Labrador & Jeong	2015	How to assess the connectivity caused by a transportation infrastructure?
39	Huang, Zhang, Zou, Yang & Li	2015	What changes are there in the environment, climate, land use and cover types for the Jiangsu Province, China?
40	Okwaraji, Webb & Edmond	2015	What are risk factors associated with access to health facilities among women in Dabat, Ethiopia?
41	Tamura, Puett, Hart, Laden & Troped	2014	How to detect spatial clusters of physical activity and obesity in relation to built-environment factors among older women in three U.S. States?
42	Shen, de Abreu e Silva & Martínez	2014	What is the impact of High-Speed Rail's on land cover change in large urban areas in Madrid between 1990 and 2006?
43	Kang, Kim & Nicholls	2014	How does the spatial distribution of domestic tourism in South Korea change between 1989 and 2011?
44	Ma, Fang, Pang & Li	2014	What is the effect of geographic proximity on scientific cooperation among Chinese cities from 1990 to 2010?
45	Zhu, Zhu, Zhang & Pan	2014	How can irrigated areas in China be mapped by using remote sensing?
46	Perchoux, Kestens, Thomas, Thierry & Chaix	2014	How do socio-demographic determinants and associations with transportation modes influence patterns of spatial behaviour?
47	Livingston & Lee	2014	How do patterns of the distribution of deprived neighbourhoods within the cities influence mortality?
48	Kulu & Washbrook	2014	To what extent are variations in fertility explained by residential context in Britain?
49	Shiode, Morita, Shiode & Okunuki	2014	What is the spatial distribution of aging populations in the Aichi Prefecture in Japan?
50	Clark, Scott &, Yiannakoulias	2014	How does the built environment and weather conditions influence the use of walking as a mode of transport?

Appendix B: Spatial Question Ontology OWL

```xml
<?xml version="1.0"?>
<rdf:RDF xmlns="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#"
    xml:base="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:xml="http://www.w3.org/XML/1998/namespace"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
   <owl:Ontology rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology">
      <owl:imports rdf:resource="http://geographicknowledge.de/vocab/AnalysisData.rdf"/>
      <rdfs:comment xml:lang="en">This ontology allows to describe the information contained within spatial
questions.</rdfs:comment>
   </owl:Ontology>

   <!--
   ///////////////////////////////////////////////////////////////////////////////////////
   //
   // Object Properties
   //
   ///////////////////////////////////////////////////////////////////////////////////////
    -->

   <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#hasGoalAttribute -->

   <owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#hasGoalAttribute">
      <rdfs:subPropertyOf rdf:resource="http://www.w3.org/2002/07/owl#topObjectProperty"/>
      <rdfs:domain
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
      <rdfs:range rdf:resource="http://geographicknowledge.de/vocab/AnalysisData.rdf#Attribute"/>
      <rdfs:comment xml:lang="en">This property defines the goal attribute of a question</rdfs:comment>
   </owl:ObjectProperty>

   <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#hasInputAttribute -->

   <owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#hasInputAttribute">
      <rdfs:subPropertyOf rdf:resource="http://www.w3.org/2002/07/owl#topObjectProperty"/>
      <rdfs:domain
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
      <rdfs:range rdf:resource="http://geographicknowledge.de/vocab/AnalysisData.rdf#Attribute"/>
      <rdfs:comment xml:lang="en">This property defines the input attribute of a question</rdfs:comment>
   </owl:ObjectProperty>

   <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#hasSpatialExtent -->

   <owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#hasSpatialExtent">
      <rdfs:subPropertyOf rdf:resource="http://www.w3.org/2002/07/owl#topObjectProperty"/>
      <rdfs:domain
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
      <rdfs:range
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#SpatialExtent"/>
      <rdfs:comment xml:lang="en">This property defines the spatial extent of a question</rdfs:comment>
   </owl:ObjectProperty>

   <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#hasSupportAttribute -->
```

```
    <owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#hasSupportAttribute"
>
        <rdfs:subPropertyOf rdf:resource="http://www.w3.org/2002/07/owl#topObjectProperty"/>
        <rdfs:domain
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
        <rdfs:range rdf:resource="http://geographicknowledge.de/vocab/AnalysisData.rdf#SupportAttribute"/>
        <rdfs:comment xml:lang="en">This property defines the support attribute of a question</rdfs:comment>
    </owl:ObjectProperty>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#hasTemporalExtent -->

    <owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#hasTemporalExtent">
        <rdfs:subPropertyOf rdf:resource="http://www.w3.org/2002/07/owl#topObjectProperty"/>
        <rdfs:domain
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
        <rdfs:range
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#TemporalExtent"/
>
        <rdfs:comment xml:lang="en">This property defines the temporal extent of a question</rdfs:comment>
    </owl:ObjectProperty>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#translatesTo -->

    <owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#translatesTo">
        <rdfs:subPropertyOf rdf:resource="http://www.w3.org/2002/07/owl#topObjectProperty"/>
        <rdfs:domain
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
        <rdfs:range
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Query"/>
        <rdfs:comment xml:lang="en">This property defines the query which the spatial question translates
to</rdfs:comment>
    </owl:ObjectProperty>

    <!--
    ///////////////////////////////////////////////////////////////////////////////////////
    //
    // Classes
    //
    ///////////////////////////////////////////////////////////////////////////////////////
     -->

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Accessibility -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Accessibility">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
        <rdfs:comment xml:lang="en">The class of accessibility analysis questions</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Accuracy -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Accuracy">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
        <rdfs:comment xml:lang="en">The class of questions which focuses on the spatial concept of
accuracy</rdfs:comment>
    </owl:Class>
```

```xml
    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Cluster -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Cluster">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#SpatialDistributio
n"/>
        <rdfs:comment xml:lang="en">The class of cluster analysis questions</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Extent -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Extent">
        <rdfs:comment xml:lang="en">The superclass of extents</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Granularity -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Granularity">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
        <rdfs:comment xml:lang="en">The class of questions which focus on the concept of granularity</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#LandUse -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#LandUse">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
        <rdfs:comment xml:lang="en">The class of questions in which land use (change/model) is a core aspect of the
analysis</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#NetworkAnalysis -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#NetworkAnalysis">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
        <rdfs:comment xml:lang="en">The class of questions which focus on network analysis</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#PopulationDistribution -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#PopulationDistributio
n">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#SpatialDistributio
n"/>
        <rdfs:comment xml:lang="en">The class of population distribution type questions</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Query -->

    <owl:Class rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Query">
        <rdfs:comment xml:lang="en">The class of queries. This is the SPARQL notation the spatial question translates
to</rdfs:comment>
```

```
    </owl:Class>


    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question">
        <rdfs:comment xml:lang="en">The class of spatial questions in Human Geography</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Regression -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Regression">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
        <rdfs:comment xml:lang="en">The class of questions which use a type of regression analysis to investigate a
spatial phenomenon</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#SpatialDistribution -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#SpatialDistribution">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
        <rdfs:comment xml:lang="en">The class of questions which analyse spatial distribution</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#SpatialExtent -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#SpatialExtent">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Extent"/>
        <owl:disjointWith
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#TemporalExtent"/
>
        <rdfs:comment xml:lang="en">The class of spatial extents. This is the study area of a question.</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#SpatiotemporalTrend -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#SpatiotemporalTrend
">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
        <rdfs:comment xml:lang="en">The class of questions which investigate a spatiotemporal trend</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#StatisticalRegion -->

    <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#StatisticalRegion">
        <rdfs:subClassOf rdf:resource="http://geographicknowledge.de/vocab/AnalysisData.rdf#SupportAttribute"/>
        <rdfs:comment xml:lang="en">The class of support attributes which are a statistical region</rdfs:comment>
    </owl:Class>

    <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Suitability -->

    <owl:Class
```

```
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Suitability">
      <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
      <rdfs:comment xml:lang="en">The class of suitability analysis questions</rdfs:comment>
   </owl:Class>

   <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#SuitabilityScore -->

   <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#SuitabilityScore">
      <rdfs:subClassOf rdf:resource="http://geographicknowledge.de/vocab/AnalysisData.rdf#MeasureAttribute"/>
      <rdfs:comment xml:lang="en">The class of suitability scores</rdfs:comment>
   </owl:Class>

   <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#TemporalExtent -->

   <owl:Class
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#TemporalExtent">
      <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Extent"/>
      <rdfs:comment xml:lang="en">The class of temporal extents. This is the study period of the
question.</rdfs:comment>
   </owl:Class>

   <!-- http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Track -->

   <owl:Class rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Track">
      <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
      <rdfs:comment xml:lang="en">The class of track analysis questions</rdfs:comment>
   </owl:Class>

   <!--
   ///////////////////////////////////////////////////////////////////////////////////////
   //
   // General axioms
   //
   ///////////////////////////////////////////////////////////////////////////////////////
    -->

   <rdf:Description>
      <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#AllDisjointClasses"/>
      <owl:members rdf:parseType="Collection">
         <rdf:Description
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Accessibility"/>
         <rdf:Description
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Cluster"/>
         <rdf:Description
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#LandUse"/>
         <rdf:Description
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Suitability"/>
         <rdf:Description
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Track"/>
      </owl:members>
   </rdf:Description>
   <rdf:Description>
      <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#AllDisjointClasses"/>
      <owl:members rdf:parseType="Collection">
         <rdf:Description
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Accessibility"/>
         <rdf:Description
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Cluster"/>
```

```
        <rdf:Description
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Suitability"/>
      </owl:members>
   </rdf:Description>
   <rdf:Description>
      <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#AllDisjointClasses"/>
      <owl:members rdf:parseType="Collection">
        <rdf:Description
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Extent"/>
        <rdf:Description
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Query"/>
        <rdf:Description
rdf:about="http://www.semanticweb.org/jgwie/ontologies/2019/1/SpatialQuestionOntology#Question"/>
      </owl:members>
   </rdf:Description>
</rdf:RDF>

<!-- Generated by the OWL API (version 4.5.6.2018-09-06T00:27:41Z) https://github.com/owlcs/owlapi -->
```