Linnéuniversitetet
Kalmar Växjö

Bachelor’s Degree Project

Ontology-Based Semantic Web Mining Challenges

A Literature Review

Author: Christopher March
Supervisor: Arianit Kurti
Semester: VT 2022
Discipline: Software Technology
Course code: 2DV50e

Linnéuniversitetet
Kalmar Växjö

Abstract

The semantic web is an extension of the current web that provides a standard structure for data representation and reasoning, allowing content to be readable for both humans and machines in a form known as ontological knowledge bases. The goal of the Semantic Web is to be used in large-scale technologies or systems such as search engines, healthcare systems, and social media platforms. Some challenges may deter further progress in the development of the Semantic Web and the associated web mining processes. In this review paper, an overview of Semantic Web mining will examine and analyze challenges with data integration, dynamic knowledge-based methods, efficiencies, and data mining algorithms regarding ontological approaches. Then, a review of recent solutions to these challenges such as clustering, classification, association rule mining, and ontological building aides that overcome the challenges will be discussed and analyzed.

Keywords

Ontology, Semantic Web mining, OWL, Ontology matching, Ontology-based association rule mining

Contents

1 Introduction

This literature review is a Bachelor thesis in computer science that focuses on reviewing literature to unveil challenges and solutions within the field of Semantic Web mining with an emphasis on ontology-based approaches. The World Wide Web we know today has most of the web content that are designed just for human understanding but not for machines. When investigating the development of the web, traditional approaches have been deemed insufficient. Web 1.0, spanning from 1995 to 2000, was characterized by a static structure primarily focused on document sharing. During this era, the web was limited to the dissemination of information without significant user interaction or dynamic content [1]. Web 2.0 (2000-2010) enabled user-generated content and data sharing, seen in platforms like YouTube, Facebook, and Wikipedia [1]. In the current web the results of searching a keyword can result in an extremely large set of listed results. The increase of web content puts more technical challenges on search engines to decide which results are more relevant or valuable to the user. Instead of human interference on the selection of results, there is an increasing need to automate the decision making process.

Compared to the traditional approaches to web development, Semantic Web is an alternative solution for decision making that was officially defined in 2010 and is currently in development [1]. Tim Berners-Lee, the creator of the World Wide Web, initially proposed the concept of the Semantic Web or Web 3.0. This revolutionary idea aimed to enhance the existing web by assigning precise definitions to information, thereby enhancing communication and compatibility between machines and humans [2].

The Semantic Web renders documents to a machine-readable format by adding semantics [3]. This means that a transformation from a document-based web to data-based web is established. However, the change has been limited due to web pages containing different forms of unformatted text or data.

1.1 Background

The semantic web represents an enhanced iteration of the present-day web, enabling improved data representation and reasoning through a more advanced structure. The data is stored using ontologies, which enable inference power over the stored data [4]. To better understand what ontologies are, imagine there exists a large collection of different kinds of beans, such as kidney, pinto, and garbanzo. The goal with these beans is to organize it in a way that makes it easy to find specific beans and determine how different beans are related to one another. One such way to accomplish this is to generate a set of rules for how the beans should be labeled and categorized. There could be categories such as kidney or pinto and within those categories there could be more categorization levels which organize the beans by color, size, or other features. The rule for organizing the beans is like an ontology. It provides a way to structure the information in a logical way to make it easy for the data miners to find and analyze specific pieces of data. Akin to how the semantic web does it, ontologies are used to structure and organize online data. We can better grasp the connections between various bits of information by using ontologies, allowing conclusions to be drawn based on that understanding. When attempting to aggregate or evaluate vast amounts of data, this can be extremely useful. Web 2.0, which constitutes the existing web, primarily serves purposes such as data searching, merging, and extraction. The efficacy and challenges linked to these tasks hinge upon how the knowledge structure is stored and represented [4]. Numerous knowledge representation schemes and languages have been developed to encompass a wide range of knowledge domains, facilitating their utilization within the Semantic Web.

As shown in Figure 1.1, the architecture of Semantic Web can be divided into seven layers: 1) URI; 2) XML, NS, XML schema; 3) RDF and RDF schema; 4) Ontology vocabulary; 5) Logic; 6) Proof; and 7) Trust. Further information about the keywords can be found in the appendix Table A.

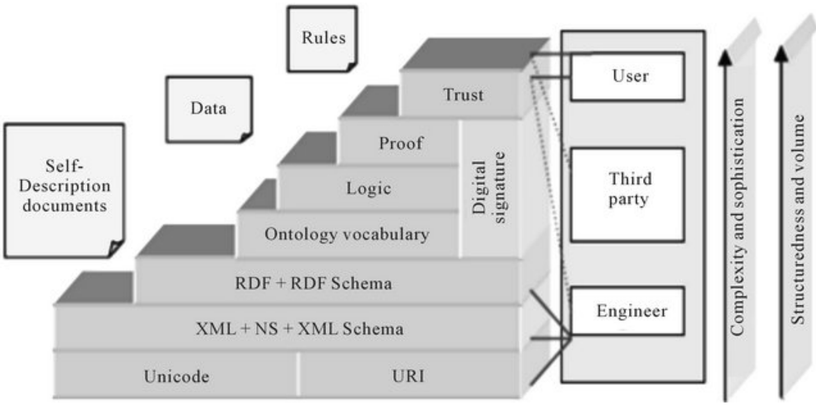

Figure 1.1: The Semantic Web layer architecture [5]

Semantic Web mining is the result of combining two fields: data mining and the Semantic Web. The main concept underlying this mining approach is to enhance the outcomes of Web Mining through the utilization of the novel semantic structures found within the Web. Advances in Web mining can also help build the Semantic Web. The process of using data mining tools to examine and glean useful insights from the content, architecture, and usage patterns of web resources is known as web mining [6]. However, there are new challenges in the jump from data mining to Semantic Web mining due to issues such as the complexity and nature of the semantic data.

1.2 Related Work

In this review, several authors were examined who conducted comprehensive surveys on the Semantic Web within the domain of Data Mining and Knowledge Discovery. Additionally, an overview of Semantic Web approaches across diverse platforms within the knowledge discovery process was explored. There is a knowledge gap for identifying the specific hurdles and obstacles in applying these techniques to web mining, even though earlier research has offered thorough assessments of Semantic Web approaches in the disciplines of Data Mining and Knowledge Discovery. A critique of semantic web mining by Yasodha [7] outlines the potential for using approaches to extract significant patterns from web activity. The researcher notes a technical problem with semantic web personalization, in which they must extract hierarchical relationships from web activity, turn them into ontologies, and then infer personal knowledge of ontologies. This creates a knowledge gap about how these ontologies might be used to classify and extract unstructured web data.

In contrast, UmaRani and Sridevi [8] present a general overview of how ontology features could be used to classify unstructured web data and provide specific use cases for ontologies employed in different web mining techniques. This emphasizes the promise of ontologies to tackle the problem of categorizing unstructured online data, but there is still a lack of knowledge about how to get past challenges like algorithmic flaws, complicated semantic data structures, and overlapped description logic.

In tandem, Quboa and Saraee [5] conducted a comprehensive survey on Semantic Web mining, referencing UmaRani and Sridevi's work to identify the challenges in applying Semantic Web mining techniques to web mining. Yet, the survey only presented generic solutions for these challenges, indicating an area for further research for current solutions.

While prior research has shown the possibility of using Semantic Web techniques for web mining, there is still a knowledge gap about the precise difficulties and barriers that must be overcome. By identifying and resolving specific barriers to using Semantic Web approaches for web mining, this lit review intends to close this gap.

1.3 Problem formulation

Numerous surveys and reviews have been conducted on Semantic Web mining practices, revealing that researchers faced similar challenges but utilized varying solutions to tackle them. However, the volume of semantic data has significantly increased as a result of the accessibility of numerous Semantic Web-based ontologies, which give data domain-specific context and semantic tags. [9]. The field of data mining is rapidly evolving, moving from mining tentative data with less background information to utilizing the vast knowledge stored in domain ontologies. The methods employed for semantic data mining, based on ontologies, aim to incorporate formal ontologies into the mining process. This paper reviews core concepts of the Semantic Web and aims to answer a couple key research questions:

> RQ1: What are recent challenges in Semantic Web mining with regards to ontology?
>
> RQ2: How can ontology be used to solve/aid the Semantic Web mining challenges?

1.4 Motivation

This literature review aims to assist researcher understand how ontologies is semantic data mining serve as a way to bridge semantic gaps between applications, data mining algorithms, data, and the resulting data mining outcomes. The research questions can help computer researchers in the Business Intelligence (BI) field to understand an overview of Semantic Web mining and ontology development challenges. Furthermore, in general, data mining is a process used by companies to transform data into useful insights. By using the results from ontology-based Semantic Web mining, businesses have the potential to acquire further insights to their clients further enabling them to create better marketing goals, boost sales, and reduce costs. Moreover, as the Semantic Web is still in development, this review will help guide business researchers in building knowledge in the field of Semantic data mining before web 3.0 becomes mainstream. These questions can also aid researchers understand recent developments with semantic web mining and how ontology development has influenced semantic web mining tasks. Finally, this literature review provides the interpretation of existing literature considering updated developments in Semantic Web mining to help with establishing consistency in knowledge of relevancy for existing materials.

1.5 Objectives

Objectives of this literature review are listed in Table 1.1

O1	Gather information on Semantic Web mining practices and ontologies.
O2	Determine recent challenges in the field of Semantic Web mining and recent work to answer the challenges.
O3	Analyze results with respect to methodology and research questions.

Table 1.1: List of literature review objectives

1.6 Scope/Limitation

The concept of the Semantic Web is very broad and contains various distinctive layers of architecture which are displayed in Figure 1.1. To better narrow the scope, this literature review will focus on the ontology-based challenges and the recent solutions to the respective challenges. This paper aims to target the unique challenges that Semantic Web mining faces when ontologies are involved instead of targeting the entire semantic architecture.

This paper is limited by the complete reliance on previously published research and the availability of these studies using the method outlined in the search methodology.

1.7 Target group

The target group for this literature review are Business Intelligence groups interested in mining outcomes for ontology based Semantic Web data mining processes. Also, this literature review targets computer science researchers who are interested in discovering how ontology mining tasks can be used to overcome Semantic mining hurdles.

1.8 Outline

This report is organized as follows. Chapter 2 discusses the methodological framework, research methods, its validity, and ethical concerns. Chapter 3 dives into a full theoretical background and discusses the ontological challenges of Semantic data mining and related concepts to better understand general Semantic concepts. Chapter 4 contains the results which showcase the results accumulated from analyzing various research articles. Chapter 5 analyzes the research gathered in the review and makes assumptions based on its impact to answer the research questions. Chapter 6 discusses the meaning, importance, and relevance of the results. Chapter 7 concludes the literature with a summary of research, and a discussion of future work.

2 Method

This literature review will attempt to answer my research questions: what are recent challenges in Semantic Web mining with regards to ontology and how can ontology be used to solve/aid the Semantic Web mining, by utilizing a systematic literature review (SLR). A systematic literature review serves as a method to identify, assess, and interpret all existing research pertinent to a specific subject area or research question. The systematic review process developed was iterative in nature and closely follows Figure 2.1 excluding step 14. Meta analysis.

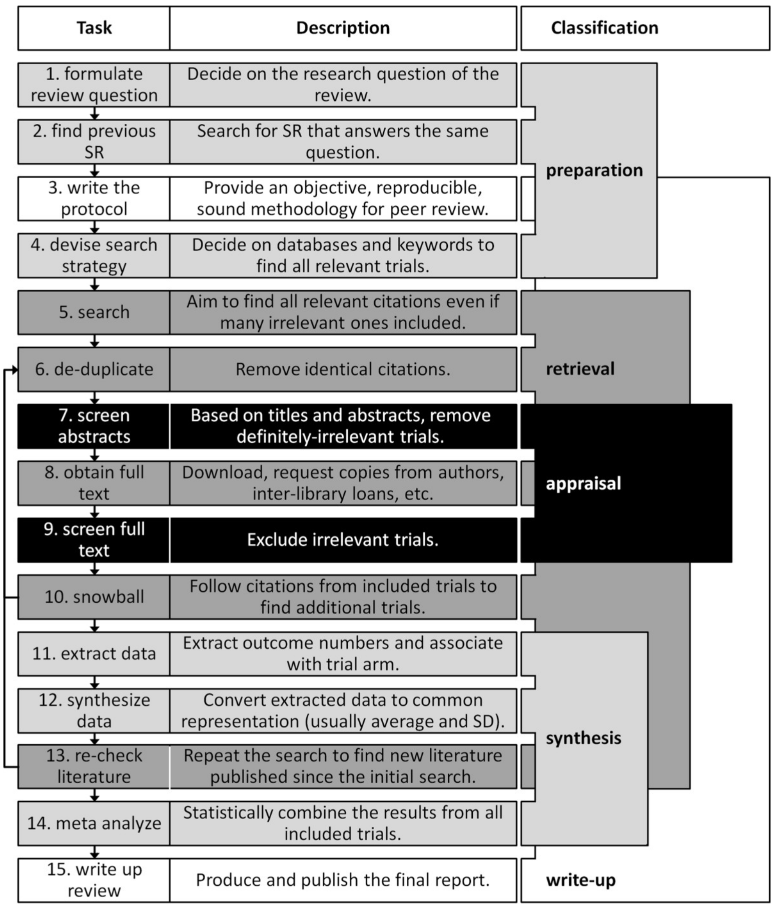

Figure 2.1: Systematic review process [10]

A study review protocol was developed for this paper to provide a clear and concise structure that is iterative in nature. This was done so the resulting literature that passes the processing checks are adequate for this literature review. In the preparation phase of the methodology the research questions are formulated which help guide the database search of Semantic Reviews (SR) that answer the same questions. Finding similar SRs allows for the basis of protocol development by providing insight into the topic along with valuable syntax and jargon. Next, the protocol can be developed by utilizing information gathered from previous SRs. The research protocol developed includes the scope, search strategy, and the presentation/organization of results. Moreover, the retrieval phase queries articles based on the search strategy developed in the protocol, additional articles are also gathered by utilizing the snowball effect which utilizes articles found via references. Articles that are found in the search phase reach the appraisal phase where the abstract and introduction is screened for relevance followed by the full text if relevancy is achieved. Finally, in the synthesis phase data in the form of literary information is extracted as specified in the protocol. The data is then compiled and summarized to answer the defined research questions which will then be organized in the final phase, the literature review write up.

2.1 Research methods

This search technique, including the databases and search words used, as well as any inclusion or exclusion criteria applied to the search results, are described in this section. It also describes the screening procedure, which entails determining which articles fit the inclusion requirements for the review by assessing the titles and abstracts of the discovered papers.

Several procedures were followed to ensure a high-quality systematic review of the literature on Semantic Web mining which were carried out utilizing guidelines loosely derived from Kitchenham [11]. By utilizing a guided review process this section aims to explain how the research questions will be answered. To start with, a thorough explanation of the search strategy is provided, together with information on the databases and search keywords that were used, as well as any inclusion and exclusion criteria that were used to filter the search results. The screening procedure is then described, including the steps taken to eliminate duplicates and determine the relevance of the remaining articles based on their titles and abstracts. After reviewing the complete texts of pertinent publications, data is then extracted and categorized. Using the proper methods, the caliber of the incorporated studies is also evaluated. Ultimately, to address the research questions, the data is processed and synthesized. The findings are then presented in a straightforward and repeatable manner.

First, a comprehensive search of peer-reviewed journals, conference papers, and reports was completed based on a few key terms including Surveys, Reviews, and Systematic Reviews. Three databases were queried including the ACM Digital Library, IEEE Xplore, and Google Scholar. Second, the reference section of each article was searched to find additional research materials. Third, key technical journals from around the world were searched independently which included the following publications: Springer Link, Journal of Web Semantics, International Journal of Computers and Applications, SCIRP Journal of Computer Science and Communication, International Journal of Emerging Trends and Technology in Computer Science, International Journal of Engineering and Computer Science, International Journal of Recent Technology and Engineering, International Journal of Intelligent Systems, Journal of Information Science, and IOSR Journal of Computer Engineering. The search process uncovered 34 peer reviewed articles published from 2011-2022.

2.1.1 Research Questions

The purpose of the research questions was to ask questions that provide a contextual challenge and meaning to the review as well as solutions to said challenges. The questions were generated by attempting to create a problem/solution context for the literature review in relation to ontology-based Semantic Web mining research. The questions were surmised into two separate research questions:

> RQ1: What are recent challenges in Semantic Web mining with regards to ontology?
>
> RQ2: How can ontology be used to solve/aid the Semantic Web mining challenges?

RQ1 was created to provide a question in the form of a challenge in the field of Semantic Web which is designed to obtain a clear understanding behind the development of the Semantic Web, as well as a core understanding of the concept of ontologies. This question also helps provide a problem for this review in which RQ2 provides a solution. RQ2 is to provide ontological solutions to Semantic Web mining tasks. This question will break down the tasks of the web mining process and extrapolate data from primary studies to provide a clear understanding of how ontology solves mining challenges.

2.1.2 Search Strategy

The main objective of this SLR was to identify as many relevant studies as possible about ontology-based Semantic Web Mining practices and ontology development and its corelated challenges. For this purpose, a search strategy was defined.

The search strategy combines an automatic and manual search approach. The automatic approach queries the three databases used in this review which were: Google Scholar, ACM Digital Library, and IEEE Xplore that were utilized to find relevant research articles [12],[13],[14]. Meanwhile, the manual approach refers to manually searching the databases for journal proceedings or conference proceedings that may have been missed from the automated search approach. This strategy utilizes the following search string for the automatic approach:

> Title:("Semantic Web Mining" OR "Semantic Web Ontology" OR "Semantic Web Mining challenge" OR "Semantic Web Mining Ontology") OR Abstract:("Semantic Web Mining" OR "Semantic Web Ontology" OR "Semantic Web Mining challenge" OR "Semantic Web Mining Ontology")

The search string was developed in an iterative process to find the most relevant articles across all three databases used in a reproducible way. The first string "Semantic Web Mining" provided many results; however, it lacked the challenges of Semantic Web mining as well as the background of the ontology language vocabulary. Therefore, the string was developed further to also include papers that matched the string queries that follow. The search string also limits articles related to Semantic Web mining practices between 2011-2022. The reason for this is to provide a SLR with a focus on recent trends with regards to challenges in Semantic ontology development and Ontology mining practices.

Often in the retrieval phase, an article containing secondary data was provided during an automated or manual search approach. These articles were parsed using the inclusion and exclusion principles, and if they met the criteria then the snowball method was executed which scans the article for related reference material that contains the primary data to be analyzed.

Due to time and resource constraints the research was constrained to options available online and free of charge. Most manual and automated search approaches were performed on google scholar and the chrome plug-in Lean Library was utilized to gain access to articles from other databases which were made available due to Linnaeus University's inter-library loan. Only publications written in English with available full text were included in this review. Additional inclusion and exclusions criteria were defined in the protocol to aid in the following appraisal phase.

2.1.3 Inclusion Criteria

Inclusion criteria was developed to apply a filter for the papers that had potential to be included in this review. This SLR included papers that passed these following criteria:

- Only studies related to Semantic Web Mining practices that were published between the year 2011 to 2022 were approved. This ten-year research period was chosen due to the increasing trend of publishing from this topic being found in the 2011-2015 date range.

- The articles must be related to the field of Semantic Web mining that addresses ontology-based approaches to mining. Also, articles which provide background and context to the concept of the Semantic Web, and ontology-based web mining were also included in this study.
- This SLR included articles that provided evidence and assessment methods of any type to validate their research claims. The validation approaches could be in the form of simulation analysis, real world examples or studies that provided opinion oriented, statistical, textual, or qualitative evaluations. For example, evidence may be in the form of experimental observations, comparisons, problem and solution, and advantages and disadvantages.
- Only papers written in the English language.

2.1.4 Exclusion Criteria

Exclusion criteria was developed to filter out papers that deemed invalid or unrelated based on the following criteria:

- The publication is not a primary study
- The full text of the article was not available for free.
- The publication is not written in English.
- The publication has a publication date prior to the year 2011.
- The article does not relate to the field of Semantic Web mining that addresses ontology-based approaches to mining or provides background or context to the research questions.
- Duplicate papers found from separate database queries.

Articles that meet the inclusion criteria were accepted for use in this review, and articles that met any exclusion criteria were removed from the search list.

2.1.5 Study Selection

The study selection process was designed to identify the relevant studies. This was achieved by dividing the process into several stages, each stage filtered and excluded irrelevant papers based on its relevance to the research questions. A visual representation of the filtering process can be seen in Figure 2.2.

In Figure 2.2 the initial query is used in the following 3 databases: Google Scholar, ACM, and IEEE Xplore. The query results stage of the process contains all resulting papers from the query search. The sorting of query results was done by relevance which further reduced the number of publications for

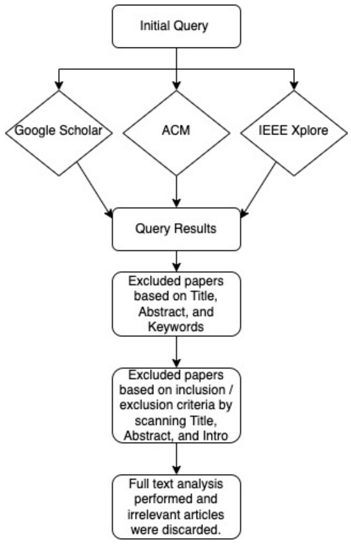

Figure 2.2: Study selection process.

further review. In the following stages each article was scanned by its title, keywords, and abstract if deemed necessary by the researcher and articles that did not relate to the research subject were extracted. The remaining publications were then further evaluated by scanning the title, abstract, and introduction to determine if the paper had further relevancy based on the selection criteria and if they don't meet the criteria then they are excluded. Finally, if the article is still deemed relevant a full text analysis was scanned for Semantic Web mining practices regarding ontology-based methods which resulted in the final set of articles being excluded. The remaining studies were selected for data extraction and used in this review.

2.1.6 Data Extraction and Synthesis

The synthesis phase concerns the execution of data extraction and data synthesis from articles that meet the criteria. Data extraction took place in the form of an annotated bibliography by utilizing a form for data extraction designed according to [11]. To gather the relevant information the form for data extraction is presented in Table 2.1.

2.2 Reliability and Validity

Using the protocol established it should be possible for other researchers to replicate the work and acquire similar results. Differences in results may occur due to the manual search approach used in Google Scholar to acquire identical results.

To mitigate bias that may have been introduced due to the research being conducted by a single researcher a randomized test was executed. Three arti-

Topic	Description
Title	Title of article
Year Published	Publication date
Publication Type	Conference article / Thesis / Journal
Valid	Peer Reviewed (True/False)
Keywords	Defined in abstract
Research Type	Survey / Case study / Experiment
Primary or Secondary	Data type
Challenges	Challenge of ontology-based Semantic Web mining
Solutions	Solutions to ontology-based Semantic Web mining challenges
Search Strategy Used	Automatic / manual approach

Table 2.1: Data extraction form.

cles that passed the final study selection phase were chosen randomly and re-evaluated by parsing them through the inclusion/exclusion criteria from chapter 2.1.4 and 2.1.3. The test resulted in all three articles being proven relevant for this review ultimately proving the repeatability of method and reliability of the review.

The review's validity is based on its purpose, scope, and audience [15]. The design of this methodology began with purpose by clearly defining and explaining the creation of the research questions. The questions also gave the SLR its scope by narrowing the focus of the research on the challenges of ontology based semantic mining challenges instead of every challenged faced in the field. The audience will find validity in the review as it utilizes standard practices of an SLR which was outlined in Chapter 2 and its subsections.

2.3 Ethical Consideration

The articles used in the SLR were all peer reviewed and collected from open source databases available for public use. Selection bias was mitigated by adhering to the protection mechanism listed in Kitchenham which utilized a quasi-randomized trial that randomized articles that passed the selection phase and parsed them through the exclusion/inclusion criteria for relevance [11]. The primary data studies were examined for possible breaches in ethical standards in the science community, therefore it is safe to assume that there are no ethical considerations to consider in this literature review.

3 Theoretical Background

This section aims to provide an understanding of data mining and ontologies to describe how they present a challenge to the Semantic Web mining. The theoretical background will aim to answer RQ1 by describing the elements of an ontology and data mining to provide a theoretical background which will support the challenges in Semantic Web mining. Furthermore, ontology-based mining techniques and tasks will be proposed to answer RQ2.

3.1 Ontology

The most important challenge of Semantic Web is ontology language and its related aspects [16]. Within the Semantic Web, ontologies play a crucial role in pushing interoperability and initiating a comprehension between various ontological entities. They serve as a fundamental component in addressing the issue of semantic heterogeneity by facilitating semantic cohesion between diverse web apps and services [17]. Ontology can be described as a hierarchical depiction encompassing classes, sub-classes, properties, and instances [18].

In Semantic Web development the ontology vocabulary is associated with various languages used in mapping multiple ontologies. During the early phases of ontology development, XML and RDF were introduced to define the syntax of content and indicate the semantics of data, respectively. Shortly after the introduction of these initial foundational languages, a more powerful and well-defined language emerged. The most common ontology language in the framework for semantic web includes OWL which was developed by the World Wide Web Consortium (W3C) which is an XML-based language for modeling and expressing ontologies [19]. OWL became a W3C recommendation in 2004 however later in 2009 OWL 2 which is a more developed version of OWL became the new W3C recommendation [19].

OWL provides machines with a better ability to understand online material as compared to other XML-based languages like RDF because of the expanded vocabulary and fundamental formal semantics of description logics (DL). DLs are logics created especially to represent and make sense of structured knowledge [19]. DLs are recognized as logical theories since they are First-Order Logic (FOL) decidable fragments.

3.2 Ontology Elements

The ontology is comprised of five main components: concepts, instances, relations, functions, and axioms [23] where:

- Concepts, also known as classes, serve as the primary formalized elements within the domain [20]. This concept is represented by a super class representing a higher class or what can be considered a parent class accompanied by a subordinate or child class.

- Instances, also referred to as objects, represent the main entities within the domain based on the ontology structure [20]. For example, the country 'Sweden' could be an instance of the class 'Scandinavian countries.'

- Relationships act as connections between concepts, serving to represent the structure of the ontology, whether it is taxonomic or non-taxonomic in nature [20]. Relationships, to put it more precisely, describe the link between one notion from the domain and another from the range. Driving, for instance, could be understood as the relationship between the domain concept of "car" and the range concept of "highway" or "road."

- Functions are components created to calculate data derived from other components [20].

- Axioms are limitations, guidelines, or definitions that control how ontology elements relate to one another. The lowest unit of knowledge in an ontology specifies the necessary logical correspondences and conditions [20]. In simpler terms, it is used to set limitations on the values assigned to classes or instances in order to ensure the consistency of the ontology.

3.3 Ontology Structure

An ontology's structure is typically described using a 5-tuple [21]

$$\text{5-tuple O:} = (\text{C, R, } H^C, H^R, \text{I}),$$

where:

- C stands for a group of ideas that are instances of the 'tdf:Class' and are part of the ontology. These ideas are arranged in a hierarchy of subsumptions to reflect their hierarchical links H^C [21].

- R is a representation of the collection of connections that bind concepts together. These linkages and associations between concepts in the ontology are defined by these relationships, which are examples of the 'rdf:Property'. $R_i \varepsilon R$ and $R_i \rightarrow$ C × C [24].

- H^C represents an instance of a binary relation that corresponds to the 'rdf:subClassOf' identifier is used to represent the hierarchy of concerns. This relationship establishes the superclass-subclass links as well as the hierarchical ties between various classes in the ontology. $H^C \subseteq CXC$, where $H^C(C_1, C_2)$ assumes C_1 is a sub-notion C_2 [21].

- H^R represents a relation to represent the hierarchy of relationships. This relation displays the ontology's hierarchical structure by representing the connections between various relations. $H^R \subseteq RXR$, where H^R (R1, R2) identifies R1 as a subrelation of R2, which is an instance of the 'rdfs:subPropertyOf' class. [21].

- I represents the instantiation of concepts within a specific domain. It serves as an instance of 'rdf:type' and signifies the classification of entities as belonging to certain concepts within the ontology [21].

The following examples provides OWL structures for a few simple specifications and properties. However, the examples make use of RDF and RDFS. RDF is the W3C standard model that is used for describing the metadata and ontology [22]. It is frequently expressed as a subject, predicate, and object (SPO) triple structure, where each triple signifies a claim or piece of information. By adding vocabularies, taxonomies, and specifying the scope of RDF classes and properties, RDFS, an extension of RDF, broadens its capabilities [22]. Owl in that regard is a further extension of both RDF and RDFS that provides the expressive language for defining ontologies that capture the semantic of domain knowledge.

- Below is an example of OWL specification given in [18]:

```
o Football and Basketball are sports.
o Football is not Cricket.
o Football is not Basketball.
o < owl: Class rdf: about = "#Football">
o <owl: disjointWith rdf: resource "#Cricket"/>
o <owl: disjointWith rdf: resource = #Basketball"/>
o </owl:Class> <owl: Class rdf: ID = "Football">
o  <owl: equivalentClass rdf: resource = "Basketball"/>
o </owl: Class>
```

- Below is an OWL specification for Property given in [18]:

```
o Gagan plays football.
o < owl: ObjectProperty rdf: about = "#plays">
o <rdfs domain rdf: resource = "#Gagan"/>
o <rdfs: range rdf: resource = "#football"/>
o </owl: ObjectProperty>
```

- Below is an OWL specification for property restriction given in [18]:

```
o Basketball is only being played by college students.
o < owl: Class rdf: about = "#"Basketball">
o <rdfs: subClassOf> <owl: Restriction>
o <owl: onProperty rdf: resource = "#played by">
o <owl: allValuesFrom rdf: resource = #collegestudents"/>
o </rdfs: subClassOf> </owl: Class>
```

3.4 Ontology Types

Various approaches exist to express or model the classification of concepts in a semantic manner. Ontologies and taxonomies are a few ways, with taxonomies being the study of classifying and arranging words in a hierarchical or tree-like structure [18]. It is specifically employed to explicitly describe concepts and the connections between them. Taxonomies and ontologies are similar, but ontologies show more complex links between ideas and properties. They also follow specific guidelines that form the framework of a knowledge base. Different kinds of ontologies are distinguished within the ontology community. While some primarily serve as taxonomies, others go further in modeling the domain and place more constraints on domain semantics [18]. These distinctions are called lightweight and heavyweight ontologies [18].

- Lightweight Ontology: Concepts and characteristics that characterize concepts are included in a lightweight ontology, which frequently presents a hierarchical structure within minor conceptual domains. Google and Yahoo are two well-known traditional search engines that have examples of lightweight ontologies [18].

- By including axioms and restrictions, a heavyweight ontology expands the ideas and characteristics of a lightweight ontology. It considers philosophical ideas in addition to the fundamental hierarchical structure to create a more intricate and semantically rich representation of the topic. With a greater understanding of the fundamental ideas and semantics, heavyweight ontologies precisely define the connections between classes and subclasses [18].

3.5 Data Mining

This section aims to introduce the concept of data mining and a few data mining techniques followed by a brief description of web mining and their respected types.

Data mining, commonly referred to as knowledge discovery from databases (KDD), is a systematic process that extracts implicit, potentially helpful, or previously unknown information from huge amounts of data [5]. To begin the data mining process, data from multiple sources is first obtained; this data is then combined to create the target data, which is a consolidated data repository. The data is pre-processed and changed in order to fit into a specified format. Data mining methods are then used to extract patterns or rules from the data. The last phase is to comprehend and produce fresh or worthwhile knowledge data utilizing these patterns and rules [5]. Data mining is crucial to the advancement of many industries, but it is particularly significant in the realm of education. It permits the identification of fascinating patterns and rules inside massive datasets that were previously unknown through the use of advanced tools like machine learning algorithms, mathematical models, and statistical methodologies [23]. There are a few widely used techniques in the field of data mining such as:

- Association Rules Mining: One of the most effective techniques for detecting frequent patterns and strong rules [24]. This particular mining technique can be characterized as an algorithm that generates patterns represented as implications of the form X $\rightarrow$ Y, where both the antecedent (X) and consequent (Y) are non-empty subsets of items. It is important to note that X and Y are distinct sets with no shared elements X $\cap$ Y= NULL [25]. Based on a variety of factors, association rule mining can be divided into distinct types. The selection criteria for item sets employed in the rule generating process define one classification. Both generalized rules and quantified association rules fall within this category. Another classification, such as horn-like rules, is based on the cardinality between X and Y [26].

- An established method of data mining that has its roots in machine learning is classification mining. It entails the assignment of every item in a data set to predefined classes or groups according to predetermined criteria [27]. Building a model or classifier that predicts categorical labels, also referred to as class label attributes, is the process of classification. It works by classifying items in a data set into particular target categories or classes. [27]. Mathematical methods like decision trees, linear programming, and statistical analysis are used in the classification method.

- A technique called clustering mining involves grouping objects according to how similar they are [28]. In order to identify relevant and previously unknown classifications or patterns in data, clustering methods are used. The process of clustering divides data into groups of related things, with different objects being placed in separate clusters. A data object may belong to a single cluster or several clusters, depending on the selected measure [28].

3.5.1 Web Mining

The web is a global network of linked files, often known as web pages, that are stored on web servers. Users can travel between pages and access a variety of information and resources on the internet because to the papers' hyperlinked connections to one another. Web mining is defined as a "quarrying (mining) of the World Wide Web (WWW) to extract useful knowledge and data about user query, content, and structure of the web" [29]. The goal of web mining is to help users find "models" or "patterns" of web page(s) that can be applied on unstructured, semi-structured, or well-formed data. Web mining can be characterized into three defined parts: Web-Structure Mining, Web-Usage Mining, and Web-Content Mining.

- Web-Content Mining: The aim of web content mining is to extract useful information or knowledge from the web page in the form of textual information. It differentiates itself from data mining because the data is sourced from mainly unorganized data structures while data mining processes mostly organized data. The web content mined from this type of mining involves text, data, image, audio, video, metadata, and hyperlinks. The objective of web content mining is to facilitate and enhance information discovery and pre-processing [30], [31].

- Web-Structure Mining: This type of mining is the method of quarrying structure information on the hyperlink structure of Web and used to improve the structure of web pages. It can be used to classify the web pages and find similarities and associations between dissimilar websites. In general, while Web-Content mining tries to find the intra-content structure (structure within a document) Web-Structure mining attempts to unveil the inter-document structure (structure within the web itself)[31].

- Web-Usage Mining: Is the collection of requests made by users to a particular website which is stored on Web server logs. This mining approach functions by extracting pertinent information from a server log, which may include user history, browser logs, user profiles, registration data, user sessions, cookies, web usage behavior, user queries, bookmark data, mouse clicks, scrolls, and more. The process of web usage mining can be divided into three distinct tasks: preprocessing, pattern discovery, and pattern analysis [29], [32].

3.6 Semantic Web Mining: Challenges (RQ1)

There exist several challenges in Semantic Web mining, however this review aims to focus on the ontological based issues that arise when researchers study various techniques to mine Semantic Web data. Reviewing various literature has unveiled challenges when applying Semantic Web mining.

Reference [26] states that the main issue in mining Semantic Web data is determining relevant transactions and items from semi-structured data which is caused by a few issues. One of the challenges is that traditional data mining techniques are designed to work with homogeneous data sets composed of transactions, where each transaction is represented by a subset of objects. On the other hand, ontology axioms that describe the conceptual domain are found in a repository of semantic annotations written in the OWL language. Assertions that are semantic annotations create associations between objects by employing characteristics that are consistent with the ontology [26]. Representing those assertions are typically through SPO triples. In this instance the identification of transactions and items increases in complexity and significance as the items may correlate with either literals or instances. Another concern arises from the fact that OWL is comprised of description logics (DLs), which are logics specifically designed to represent and reason about structured knowledge [19]. Thus, instances belonging to the same OWL class may exhibit diverse structures, resulting in issues related to structural heterogeneity [26].

Reference [33] emphasizes that an issue of the Semantic Web ontology structure happens when a traditional decision tree algorithm is applied which makes it difficult to use the semantics of the ontology. Due to the characteristics of the Semantic Web ontology, data used in decision trees are separated into classes, and these classes are related to each other in various ways like a subclass and property relationships [33]. Furthermore, when using this mining algorithm, complexity develops because semantic data's ontology network structure permits an infinite number of properties to describe a resource.

The work done in [34] mines RDF on a statement level based on the SPO view of RDF data. Any of the portions of an RDF statement can be chosen as the focus for mining by using the context of the statement as a grouping criterion. Non-frequent subjects were found to be difficult to mine when using RDF data, according to the study. Each subject appears as often as the predicates that determine its value. Similarly, there are roughly 100 times more distinct objects than predicates [34]. This could lead to a reduction in the support threshold of the overall mining process thus producing a substantial number of irrelevant patterns.

3.6.1 Ontology Challenges

Since ontologies can be considered the core of Semantic Web mining tasks, the most important challenge of the Semantic Web is ontology and its related aspects. Ontology design is still in need of improvement and needs to be evaluated in light of actual applications since it serves as the Semantic Web's knowledge representation language [16]. Thus, some ontology related development issues that should be addressed such as: ontology matching and ontology integration.

- Finding correspondences between semantically related ontology entities is known as ontology matching [35]. Simply put, the matching procedure seeks to create an alignment A for the ontologies O1 and O2 in question. The matching challenge therefore entails locating an alignment between these ontologies when two ontologies are given, even if they are simple and each contain just one entity. [36]. However, there are several challenges that face the development of ontology matching. One such challenge is the efficiency of matching techniques, because the execution time of ontology matching uses a significant allocation of main memory or bandwidth, alongside the consideration of other computational resources such as the CPU [35]. Therefore, the challenge is to develop a scalable ontology matching reference solution. In ontology matching, handling background knowledge is another difficulty. Ontologies are created in a context that incorporates previous knowledge, but this background knowledge is not always incorporated into the final ontology specification, which presents a challenge. Adding context can bring additional information that enhances memory, but it can also introduce false matches and impair precision, making it difficult to strike a balance [35].

- Ontology Integration: involves the ability to create a new ontology that is more comprehensive or appropriate for a particular application which involves combining or reusing existing ontologies. In order to create a new ontology that is more comprehensive or general, input ontologies must be combined or merged [19]. Ontology integration is a special process that involves changing one of the input ontologies (known as the target ontology) while leaving the source ontologies unaltered. Due to the enormous volume of data involved, integrating ontologies in the context of Big Data is a difficult task. Furthermore, heterogeneity problems are inevitably brought on by the existence of numerous conflicting, incomplete, and overlapping ontologies in a given domain. Ontology integration is still a crucial part of ontology development tasks despite these difficulties because creating ontologies from scratch takes a lot of time and money [19].

3.7 Ontology-based Semantic Web Mining (RQ2)

To assist in various data mining tasks, various studies leverage the use of ontologies with their formally encoded semantics. This section aims to summarize some important data mining techniques and algorithms to unveil work that has been done to successfully use ontologies for semantic data mining.

3.7.1 Ontology-Based Clustering

As mentioned in section 3.5 clustering is data mining technique that organized members into groups that are similar in some way [28]. The work done in [37] presents a Website Key Objects (WSKO) extraction technique. The term "WSKO" refers to components of a website that provide end users with useful content and formatted information. These components improve the user experience by offering interesting and pertinent information [37]. Without regard to an object's original format, the extraction of WSKO was established as a fundamental ontology that offered a standardized framework for categorizing each one. This ontology made it possible to compare objects in pairs without taking into account their original formats [37]. It was necessary to perform a preliminary ontology learning step before using the clustering algorithm. The core ontology used in this study was created using a representation of the website, including all of its web pages, web objects, and web concepts related to each object [37]. From there, the clustering algorithm compared objects with their respective concepts and all objects are grouped together according to their relevance for the end user.

The work done in [38] declares that ontologies can be used to solve text clustering challenges. Text clustering suffers from polysemy and synonymy which are terms that map to multiple meanings or the same concept as different words in a document respectfully [38]. By leveraging ontologies this research has discovered that it can reduce the number of features needed in text clustering tasks. Using an information gain measure to determine the contribution of each cluster-aware noun in a text corpus to clustering allowed a 90% reduction in the number of features from this approach to capture the main themes of the text corpus [38].

3.7.2 Ontology-based Association Rule Mining

A method for finding patterns or reliable rules is association rule mining [24]. Work done in [39], a Semantic Web Association Rule Mining (SWARM) strategy is introduced with the goal of producing Semantic Association Rules from RDF ontology data. The objective is to create a connection between the instance-level and schema-level to add more semantics to the rules [39]. This method automatically pre-processes RDF data by generating a 2-tuple semantic item related to the tuple and generates common behavior sets based on the semantic items. Finally, the pre-processed information is mined for semantic association rules by extracting knowledge encoded in the ontology.

A unique strategy for extracting association rules from heterogeneous semantic data repositories defined in RDF(S) and OWL is proposed by the research published in [26]. The technique was created to take advantage of the ontology's schema-level knowledge by extracting and combining instances of interest (features) from a medical repository and turning them into transactions

[26]. It works by having a user specify the kinds of patterns they are interested in returning from the repository by extending SPARQL (RDF query language) with new statements that allow specification of a mining pattern. Then a transaction extractor is used to construct transactions according to the users mining pattern The instance transactions will be generated by computing the composition triples involved in the mining pattern which are then fed to the association mining algorithm to produce the rules [26].

3.7.3 Ontology-based Classification

This data mining task aims to categorize each item in a dataset into a predefined set of classes or groups [27].

Research proposed by Allahyari et al. [40] indicates that text categorization can be accomplished by utilizing the semantic correspondences between a text document's content and a pertinent area of the ontology. For automatically categorizing text documents into dynamic topic categories of interest, they suggest an ontology-based solution. The study converted Wikipedia into an RDF ontology using a modified DBpedial tool which facilitated a better discovery of named entities in the document. They employ the HITS algorithm [41] to build a semantic graph comprising interconnected entities. This graph is then utilized to identify the central entities that play a significant role in the dynamic topic identification process. Finally, the classification of a document into the defined ontological context (topics) was based on calculating similarities of the documents thematical graph to each of the defined contexts [40]

3.7.4 Semantic Decision Tree

As stated in section 3.6, one of the issues faced with semantic web mining regarding ontology is the possibility of having an unlimited number of properties which describe a resource. Research in [36] have developed modified decision tree called the Semantic Decision Tree (SDT) to overcome these limits. The SDT starts by setting a target class and a target property, the target class contains target instances while the target property is one of the properties that is a datatype property whose range type is Boolean. Thus, target instances which belong to target class have true or false value. The SDT algorithm starts with a single root node which contains every instance from target class, the learning process split the root node to move on to find candidate refinements around a center class [33]. The candidate refinements search process algorithm explores the ontology after defining a center class to find information on whether the center class contains subclasses and property values whose domain is the center class or not. By adding extra information about the connections between concepts and the functions of objects within the ontology during the mining process, which is referred to as named properties in OWL, additional improvements were made. Also, since Semantic Web ontology is too large to select variables for decision trees by person, they modified the SDT to select variables based on relationship information of ontology and statistical basis [33].

3.7.5 Data Mining Ontology

In the field of ontology development there is no universally established data mining ontology yet. However, there are several ontologies currently in their developmental, one such noteworthy ontology is Data Mining Optimization Ontology (DMOP) [45]. DMOP provides a thorough conceptual framework that makes it easier to analyze various data mining tasks, algorithms, models, datasets, workflows, and performance metrics. By establishing significant connections and relationships between these components, it offers a consistent methodology for data mining analysis. The main goal of this ontology is to support meta-mining of data mining experiments to extract workflow patterns.

Semantic meta-mining is set apart from traditional meat learning by following a few basic properties. The first factor that affects it is the understanding of the data mining process and its elements, which is reflected in the ontology and knowledge base for data mining [42]. Finally, it explicitly investigates data mining algorithms from numerous angles to create a link between the features of the data and the data mining algorithms and the observed performance of learned hypotheses [42]. To support meta-mining DMOP hold the detailed taxonomy of algorithms used in data mining processes which allow meta-miners to generalize over algorithms and their properties.

4 Results

The literature review search was carried out according to the search strategy explained in chapter 2.1.2 resulting in 113,000 articles. Applying the study selection in 2.1.5 further filtered the number of articles used in the SLR to 16. Figure 4.1 displays the steps performed during the study selection process as well as the number of papers that satisfied the selection criteria. There were 34

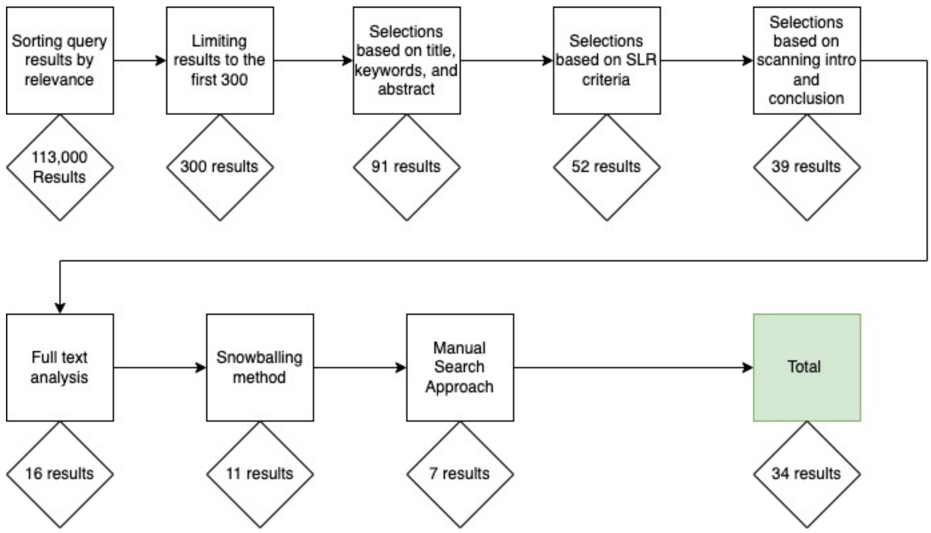

Figure 4.1: Publication selection process

total studies selected for the literature review of which 28 were primary studies that in part solved or provided context to the research questions in section 2.1.1 regarding Semantic Web Mining challenges and solutions. Of those 28 primary studies, 18 of them were obtained from the automatic search approach while 10 of them were attained from the manual search approach and included in the SLR. The full list of studies selected for use in the SLR is presented in Appendix 1.

The research types from the final selection of papers used in the SLR are presented in Figure 4.2. They consisted of 3 books, 1 Master's thesis, 3 Technical Reports, 9 conference papers, and 18 articles. The most common research methods used in the chosen studies were prototype, case studies, and qualitative studies which is displayed in Figure 4.3.

Studies in this SLR were published between 2011 and 2022, with most studies from 2011. The spread of studies along the timeline is shown in Figure 4.4. Most studies were published in 2011and 2013, after 2013 there is a clear decreasing trend in publications in the field of Semantic Web mining. However, since this SLR collected a small assortment of search results, the results are not indicative of the overall trend of Semantic Web mining research.

The following concepts were discovered to have an impact on Semantic

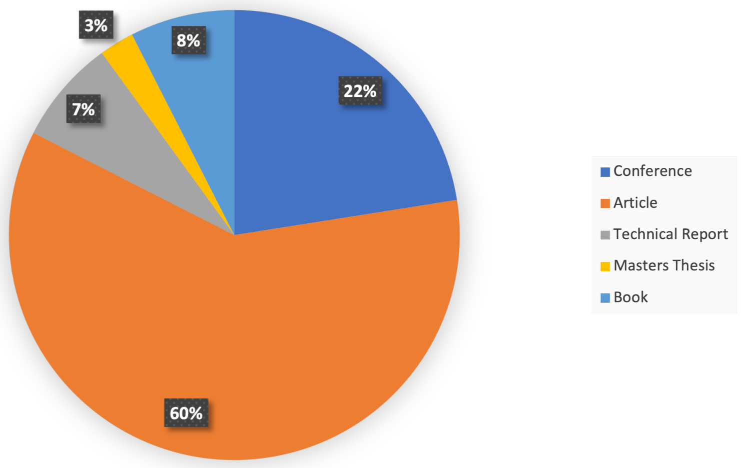

Figure 4.2: Publication Types.

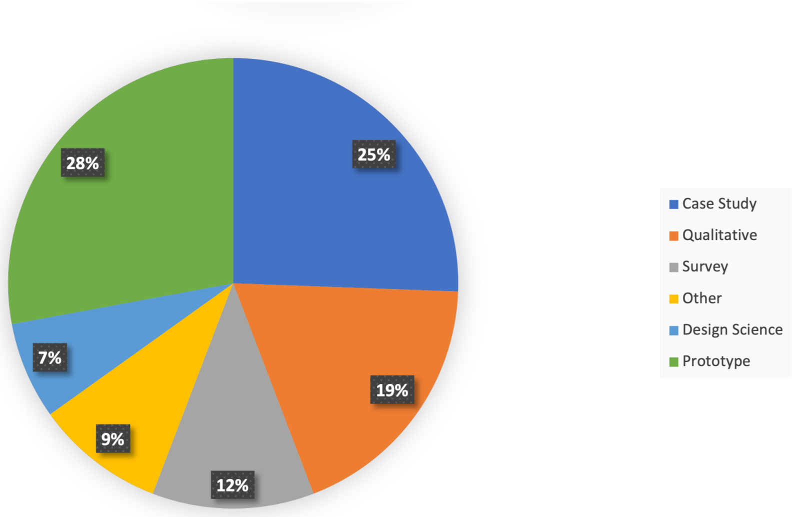

Figure 4.3: Publication Research Methods

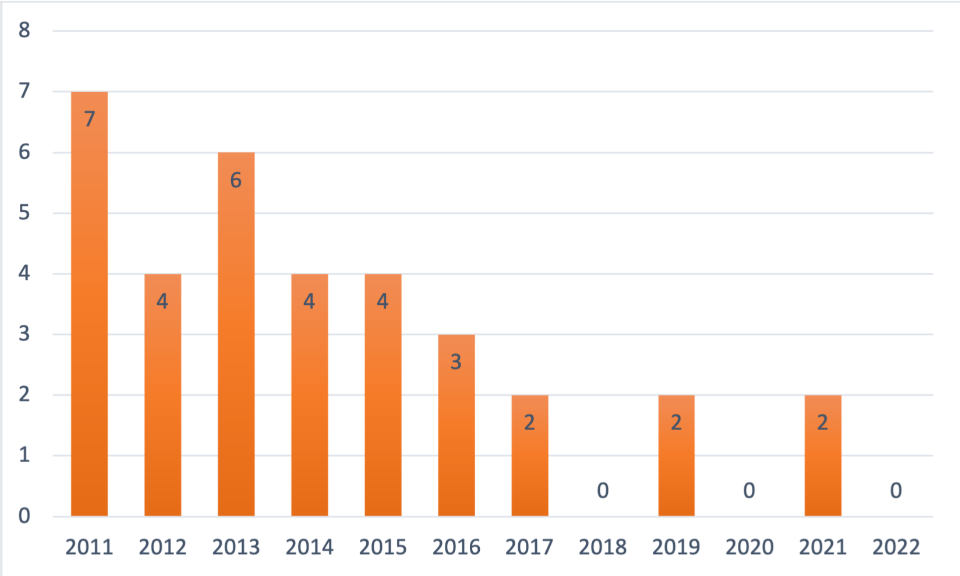

Figure 4.4: Published years of reviewed papers.

Web mining practices. The RDF is found in many papers due to it being the core of the OWL ontology language which could suggest that most research on Semantic data mining deals with the concept of RDF tuples. Ontology heterogeneity concept suggest that the issue of heterogeneous data held in semantics of ontologies is of concern to the mining practices researched. Association Rule mining appears in many studies because there are many ontological challenges and researched approaches to mine the rules from the Semantic Web using various algorithms. Clustering and its related Classification practices concepts present both challenges and solutions to Semantic Web mining by grouping objects in the same cluster as each other and discovering a model or function that accurately describes and distinguishes different data classes or concepts. The number of studies that address these concepts is shown in Table 4.1.

Concept	Number of studies
Resource Description Framework (RDF)	25
Ontology heterogeneity	20
Association Rule Mining	15
Clustering Ontology practices	12
Classification Ontology practices	13

Table 4.1: Concepts impacting Semantic Web mining.

Based on keywords of the papers used in this SLR, a word cloud was made to represent the main concepts found in research. The word cloud is shown in Figure 4.5.

Figure 4.5: Keyword and concepts cloud

5 Analysis

This section provides further context of the research questions defined in Chapter 1 and will serve as an extension of findings provided in sections 3.6 and 3.7 by analyzing the quality of the research reviewed.

5.1 Semantic Web Mining Challenges (RQ1)

There were several challenges identified in the research related to Semantic Web mining. The distribution of challenges in the primary studies is provided in Figure 5.1.

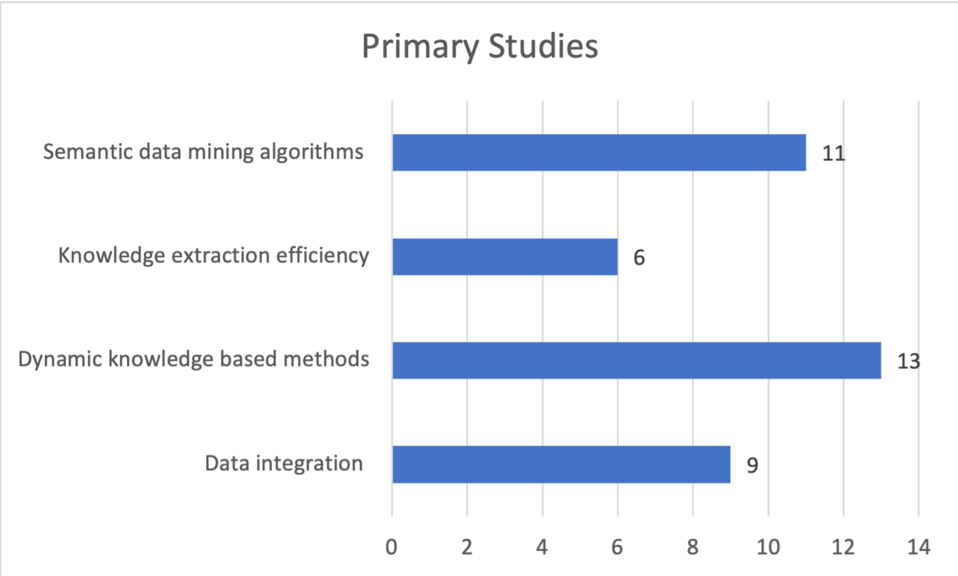

Figure 5.1: Primary study distribution regarding RQ1

5.1.1 Semantic data mining algorithms

When analyzing the primary studies of mining knowledge from the Semantic Web, a common challenge of algorithm usage and deployment was apparent. Traditional data mining techniques were largely created for homogeneous data sets, which presents a difficulty when using them to mine semantic data [S26]. Thus, most algorithms being tested in the primary studies required adjustment which causes additional challenges such as lack of automation in the data mining process and erroneous and hidden data results [S2], [S3], [S6], [S7], [S13], [S14], [S18].

The clustering technique can be used to find and establish a set of foundational classes derived from the data when an ontology is being created without the assistance of a domain expert. However, the issue with this technique is that the accompanied algorithm is distance-based which describes clusters by enumerating their members or another arbitrary way which ultimately does not provide useful data for naming the detected clusters [S3]. Furthermore, data mining has primarily focused on mining instance-level data. However, in the Semantic Web, knowledge exists at both the instance-level and the schema-level. Ignoring the schema-level can negatively impact the results of the mining process [S3], [S6], [S7], [S13], [S15], [S24].

Ontology integration/matching is another such data mining preprocess that takes two ontologies and merges them into one while preserving the ontologies coherence and knowledge. Ontology matching and integration can improve scalability and interoperability in the Semantic Web context. There are a few systems that can handle multiple ontologies in a single invocation, but the majority of existing ontology integrating algorithms are only intended to integrate a pair of ontologies. This is due to the pairwise ontology matching algorithm gaining more popularity compared to the holistic matching approach [S2].

5.1.2 Knowledge extraction efficiency

Mining the semantic web ontologies provides data scientists with better results to its respective domain, can provide new acumen from the semantic annotations, and solves complex problems from heterogeneous data to improve results from web mining. Thus another challenge is the need to extract information and knowledge efficiently and effectively. Efficiency is a performance evaluation criterion; therefore, the runtime of semantic web mining algorithms should be comparable to that of existing algorithms [S26]. To extract knowledge from semantic web data efficiently then a machine learning algorithm is needed [S14].

One such algorithm is the decision tree which is considered one of the most popular classification data mining algorithms and has gained popularity for its efficient performance on large size of data sets and interpretable representation of the results that are discovered. However, due to the complexities derived from the heterogeneous nature of semantic web ontologies it is difficult and inefficient to apply traditional decision tree algorithms and take advantage of the semantics of the ontology [S3], [S8], [S14], [S18].

5.1.3 Dynamic knowledge-based methods

Yet another challenge is the fact that ontologies are not able to express everything in the real world. Information from an ontology model can represent only a small portion of a domain competently. Having accurate information is essential for proper reasoning in computer applications. Thus, the capability of the data mining technique is bound to the completeness of the knowledge in the ontology model used in the data mining endeavor. To solve this problem the ontology must be made dynamic and adjustable without human intervention. The process of defining domain concepts and their relationships, which results in the development of an ontology, is challenging and frequently calls for the assistance of domain specialists. This job, which demands a high level of skill, frequently calls for an expert. [S3]. Ontology learning, also called 'ontology engineering', or 'ontology generation' is the semi-automatic creation of ontologies. New technologies such as ANN (Artificial Neural Network) made for web information source detection and CI, which stands for Computational Intelligence, encompasses a collection of techniques primarily centered around evolutionary computation and Artificial Neural Networks (ANNs). These methods are employed to tackle complex problems and enable machines to learn and make intelligent decisions and have been implemented to automatically construct ontologies from the current web. The major drawback from using these technologies is that they focus on solving problems separately which is ill equipped to deal with the ambiguity, autonomy, inconsistency, and uncertainty of the Semantic Web [S27].

In Ontology integration the researchers rely on tools to integrate two ontologies, however the current tools for accurate integration results are usually semi-automated which directly affect the scalability meaning that the tool is inadequate for large data sets from the semantic web. Additionally, since ontology integration tools are typically semi-automated, human intervention is required to address any problems or redundancies that may come up during or after the creation of the integrated ontology [S2].

5.1.4 Data Integration

Also, big data comes in a variety of data sources and format, making it difficult to handle with the standard data-mining techniques. OWL is comprised of DLs which are knowledge representation formalisms with understood semantics and formalisms. Thus, annotated data will not contain a rigid structure where the same OWL class may have different structures giving place to structural heterogeneity issues [S13]. The semi-structured and heterogeneous nature of Semantic Web data makes it difficult to find relevant transactions and items when working with it. Utilizing the knowledge that ontologies provide becomes essential in this situation [S2], [S7], [S13], [S16], [S22]. To harness their capabilities, ontologies must address the challenge of semantic heterogeneity by effectively integrating their distributed knowledge. However, integrating ontologies in big data is considered impossible due to an extensive amount of data sets and their associated complexities. Thus, it is strongly desired in the field of data mining research [2].

Overcoming semantic heterogeneity problems could be solved by a technique called ontology matching [S16]. However, the field of ontology matching has its own challenges that it's facing with a decrease in research progress in recent years. An ontology matching challenge was established where 3 separate databases WordNet (lexical English database), DBPedia (collection of things tied to the English language) and GTAA (a Dutch thesaurus used to index TV programs) were involved in matching with 10k, 100k, and 1 million entities respectively per ontology to be designed and conducted. The issue with this challenge is the due to many heterogeneous entities from the different domains must be matched, thus requiring an automated matching tool. However, to test for ontology matching accuracy, a reference alignment against the automated results must be created which requires manual building techniques that proved to be too demanding thus further adding another challenge to the field of ontology matching [S16].

5.2 Ontological solutions to web mining (RQ2)

Several solutions were discovered which solved some of the challenges that affected Semantic Web mining which can be found in Figure 5.2.

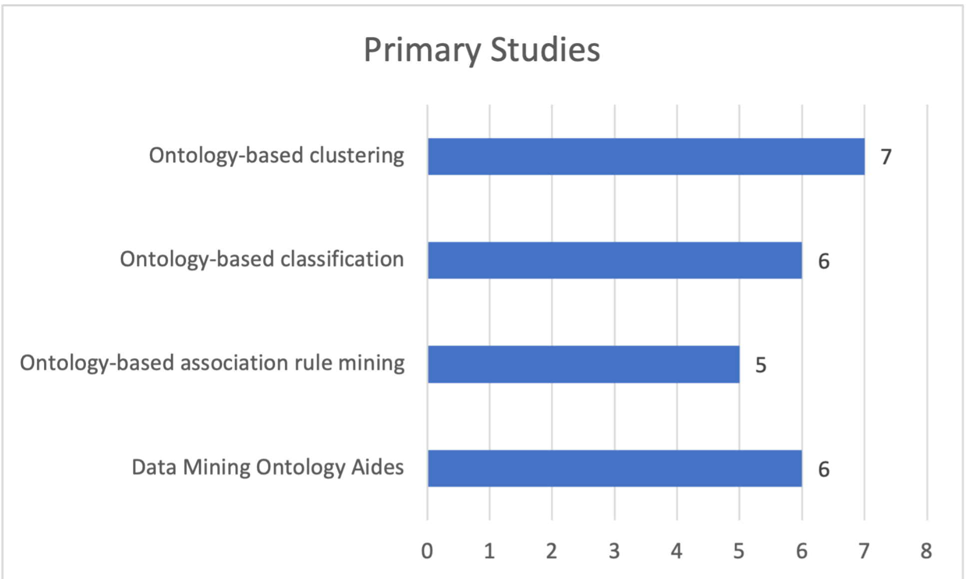

Figure 5.2: Primary studies regarding RQ2.

5.2.1 Ontology-based clustering

Utilizing clustering techniques, which entail determining similarities among data objects and grouping them into clusters based on their proximity, is one potential method for overcoming difficulties in ontological web mining. This method's goal is to automatically identify useful classes or categories within a given dataset without human supervision. A foundational set of classes can be created using clustering in the absence of a domain expert based on the patterns seen in the data. It should be noted, though, that some clustering algorithms rely on distance measurements and might not give the clusters they find explicit labels or names [S3], [S6], [S9].

One study develops a method to mine website keywords that define the search process for a group of users within knowledge discovery databases using different clustering techniques. Two different clustering techniques Self Organizing Feature Maps (SOFM) and k-Means, both considered machine learning clustering algorithms, were to be run on the same data sets to determine if both techniques returned similar results in the semantic database. The results proved similar for both algorithms and were further verified by comparing the algorithm results with a survey taken by a control group of users that were shown pages containing key objects and recording their receptivity towards it. The results displayed an 80% match between survey results and the algorithm results [S18]. Ontologies may be used to reduce the amount of features needed for a document clustering task, according to another study. It is feasible to portray documents in a more condensed and understandable way by utilizing the hierarchical structure and semantic links represented in an ontology. By concentrating on the most important parts of the data, this decrease in feature space can result in better efficiency and efficacy in clustering algorithms. Ontologies can therefore be incorporated into document clustering tasks to improve performance and enable more precise and intelligent analysis of text documents. For example, by utilizing ontology a set of semantic features for each text was identified. The features can be used for clustering once they have been found. This clustering method has the potential to produce clusters that accurately represent the text's underlying themes while also drastically reducing the number of features—often by 90% or more. [S24].

5.2.2 Ontology-based classification

Each item in a data set is divided into preset classes or groups using a machine learning-based data mining approach. By taking into account a vector of qualities connected to each item, this classification is accomplished. To produce predictions about the class labels of things that are not yet visible, the machine learning algorithm discovers patterns and relationships within the data. The method assigns items to the relevant class or group by taking use of their qualities, thereby categorizing and organizing the data. When working with complex data sets where manual classification may be time-consuming or unworkable, this method is especially helpful [S3]. Decision tree algorithms is one of the most popular classification data mining algorithms, which has gained its popularity due to its efficient performance on large size of data sets and interpretable representation of results. However, the traditional decision trees are incapable of handling the complexities and semantics of the Semantic Web. Therefore, an algorithm was proposed in [S14] which can perform the decision tree algorithm on the semantic web-based ontology. Thus, a solution was the development of the semantic decision tree which was developed with some core characteristics such as relationship information that is composed of object property and class of the ontology should be included, the use of a description logic-based constructor to expand conditional power, and an automatic variable selection method to handle the large amount of data from semantic web ontologies [S14]. An experiment was carried out to test the validity of the semantic decision tree where researchers applied the algorithm to predict the heading direction of trains and get the definition of the eastbound trains. The experiment was only able to identify 2 of the 5 eastbound trains due to a false value of target property in their algorithm. Furthermore, the experiment required a purpose-built ontology to represent the trains and the number of cars each train was pulling. Overall, the experiment proved that the semantic decision tree required additional ontology refinement for the train description logic and the algorithm needed additional fine tuning to extract the correct knowledge inferred in the ontology [S14].

Furthermore, Text documents can now be automatically categorized into dynamic subjects of interest using ontology-based techniques. These techniques enable quick and precise classification by taking advantage of semantic similarities between document content and ontology concepts. By utilizing DBpedia ontology, entities are identified from the text document and a semantic graph was constructed from the set of relations. Documents were classified by comparing their semantic graphs to determine ontological topics [S24].

5.2.3 Ontology-based association rule mining

In various stages of the mining process, including task design, data understanding, result dissemination, and result interpretation, researchers have developed association mining tools that incorporate ontologies. [S24]. Yet another solution to RQ1, association rule mining is widely recognized as a fundamental data mining task and finds applications in various domains [S13].

Further research has been done in mining semantic association rules from RDF data. By utilizing schema-level knowledge, the unique approach known as Semantic Web Association Rule Mining (SWARM) which was developed to embed semantics into rules. SWARM utilizes the rdf:type and rdfs:subClassOf relations at the schema-level to automate the process of mining semantic association rules from RDF-style knowledge sources. This makes it possible to generate rules that are enhanced with semantic data [S6]. However, Semantic Web data suffers from the lack of correctness and consistency between entities at the instance-level. Ontology definition inconsistencies and underlying data may lead to erroneous data and interpretations.

5.2.4 Data mining ontology aides

Lastly, an important tool developed to support decision-making steps that determine the outcome of data mining processes is the Data Mining OPtimization Ontology (DMOP) [S22]. One of the most important uses is to inform manual selection of various entities such as algorithms, parameters, and models. DMOP can be used by data miners unfamiliar with ontology tools as well as experts alike. For the undisciplined data miner DMOP contains a user-friendly predesigned template to populate areas of the ontology with stable concept and property definitions. For the data mining experts, they can develop the assigned module using their preferred ontology editor and submit it in the form of an OWL file. After validation, the module becomes an integral part of the ontology [S22]. However, the downfall with DMOP is when it comes to modeling data. DMOP suffers from relating instances to classes and using classes as instances, finding, and resolving undesirable deductions caused by property chains, representing attributes where the solution is ontology-driven yet merged with OWLs built in data types to aid their reuse in other applications, and linking to a foundational ontology. However, most of these issues have been resolved by utilizing features developed in the OWL 2 ontology resulting in highly axiomatized and complex ontologies [S22].

A system called SAMBO was created at Linkoping University with the express purpose of matching and combining biomedical ontologies. It gives 1:1 alignment results for concepts and relations and supports ontologies represented in OWL format. The system is very user heavy with a user having to define weighted values to ontologies for the matching to occur, however an expert in data mining would find the system useful as it checks in with user validation on the matches for additional feedback in later steps [S16].

6 Discussion

In this section, I will discuss the findings of my study on the challenges in Semantic Web mining with regards to ontology, as well as how ontology can be used to aid in solving these challenges. Through a review of various literature, I found that there are several issues that arise when researchers attempt to mine Semantic Web data using ontological-based techniques. I also found possible solutions through the use of ontological based mining techniques and applications.

While answering RQ1 several recent challenges were identified in the research papers that met the search methodology. I noticed that one of the main challenges identified in the literature is the difficulty in determining relevant transactions and items from semi-structured data. This is due to the fact that Semantic Web data is heterogeneous and calls for specific methods, whereas traditional data mining algorithms are made for homogenous datasets [26]. However, a solution with regards to RQ2 exists as ontology axioms define the conceptual domain in a repository of semantic annotations written in the OWL language, and semantic annotations are expressed as assertions that are consistent with the ontology. This makes the identification of transactions and items more complex and significant, as the items may correlate with either literals or instances [26].

Another challenge that I identified from the literature is the issue of structural heterogeneity, which arises from the fact that instances from the same OWL class may have various structures. This posed a problem when traditional decision tree algorithms were applied, as the ontology's semantics cannot be utilized effectively [33]. The complexity of the network structure of ontologies in the semantic data, which may have an unlimited number of properties describing a resource, only adds to the challenge. One ontological solution was an algorithm proposed in [33] which can perform the decision tree algorithm on the semantic web-based ontology. I believe that this recent solution is beneficial because it allows for automatic variable selection to handle the large heterogeneous data from semantic web ontologies.

I found that ontology and its related aspects are considered the core of Semantic Web mining tasks and they contain some important challenges to the research of Semantic Web mining. In particular, ontology matching and ontology integration are two ontology-related development issues that I propose need more attention in future research. I also found that ontology matching faces several challenges, such as the efficiency of matching techniques, matching with background knowledge, and others [35]. Ontology integration, on the other hand, involves the process of reusing or unifying existing ontologies to build a new, more general or more complete one that can be used by input ontologies that were integrated or merged [19].

Despite these challenges I have exposed, there are various data mining techniques and algorithms that leverage the use of ontologies with their formally encoded semantics to assist in various data mining tasks. For instance, the study in [17] proposed a novel approach for ontology-based association rule mining, while the work in [18] presented a framework for discovering patterns in RDF data that incorporates the semantics of the ontologies. These and other studies show the potential of ontology-based techniques in addressing the challenges of Semantic Web mining (RQ2).

In conducting this SLR one of my main concerns were my experience limitations, as I have never completed a project of such complexity before. Also, another problem I faced was that the many descriptive details of machine learning algorithms were omitted due to the complexities of the theories that were used to develop them which forced me to research a large amount of materials and articles that fell outside the scope of the research questions. Furthermore, I omitted data mining techniques that did not capitalize on the use of ontologies in the mining stages in order to align the goals of this literature review with the research questions. Resulting in limiting the scope of this literature review to ontology-based challenges and solutions.

To conclude my discussion, my study has identified several challenges in Semantic Web mining with regards to ontology and discussed how ontology can be used to aid in solving these challenges. The findings of my study are in line with related work in the literature, which have identified similar challenges and proposed various solutions to address them. However, a challenge still remains due to the rapidly evolving field and a growing number of new research articles.

7 Conclusions and Future Work

The objective of this literature review was to identify ongoing and current challenges in semantic web mining with regards to ontological approaches, and to provide recent solutions that tackle these challenges (RQ1 and RQ2 respectively). Through a systematic literature review, I have gained insight into the field of semantic web mining, which in turn has allowed me to evaluate research gaps within the data mining community, by analyzing challenges and solutions within the field.

To answer research question 1: What are recent challenges in Semantic Web mining with regards to ontology? In my research I found the following recent challenges within the field of semantic web mining: Data mining algorithms, knowledge extraction efficiency, dynamic knowledge based methods, and data integration. For example, data mining algorithms struggle to handle the rich semantic information contained in ontologies, making it difficult to extract useful knowledge. Knowledge extraction efficiency is also a challenge, as traditional data mining algorithms may not effectively leverage the semantic relationships between concepts in the ontology. Dynamic knowledge-based methods, which incorporate feedback from users and other sources of information, can improve the accuracy and relevance of extracted knowledge. Data integration poses yet another challenge, as data from multiple sources with different ontologies and schemas must be combined. To address these challenges, I propose that new algorithms and techniques are needed that can effectively leverage the semantic information in ontologies and improve the efficiency and accuracy of knowledge extraction and data integration.

The research indicates that automated ways to extract data and variables from semantic ontologies are helping data mining experts to answer research question 2: How can ontology be used to solve/aid the Semantic Web mining challenges? This literature review uncovered several solutions to the current challenges in semantic web mining. For example, ontological clustering, ontological classification, ontological association rule mining, and ontological data mining aides. Ontological clustering is a technique for ontological automation, which helps aid the challenge of data integration by finding similarities in data objects and placing the most similar ones into a common ontological cluster. Furthermore, techniques for ontological classification are used to pinpoint the fundamental elements of the semantic web and make it easier to pinpoint dynamic subjects. For example, a Semantic Decision Tree algorithm was developed so that it could handle the rich semantic information contained in ontologies, which solves the challenge of data mining algorithm shortfalls and knowledge extraction efficiencies thanks to its improved performance on large data sets. Moreover, ontological association rule mining solves the issue of tying instance-level data to schema level in order to attach additional semantics to pattern finding rules. This solution helps to answer the issue of dynamic knowledge based methods by taking advantage of rdf-types and rdfs-subclass relations to generate semantically enriched mining rules. Applying these ontological techniques allows the semantic web to be mined by special algorithms, which allows researchers and data miners to overcome the recent challenges of Semantic Web mining.

Further research could be performed on the advancements of machine learning algorithms and how they are progressing the field of data mining, as well as how they are currently being adapted to extract knowledge from semantic ontologies and knowledge bases. Additionally, more research could be done to determine what steps are being taken to further automate the various steps of the data mining process, allowing for greater accessibility to the field of data mining for non-domain experts. Lastly, it is important to note that new challenges to the field of semantic data mining are ever-evolving due to rapid development of this niche field.

In summary, this literature review has outlined the ontology-related obstacles that occur in semantic web mining, as well as the solutions that have been proposed. Future research points to continued improvements in automated methods for extracting data from semantic ontologies as the field advances, as well as the creation of fresh approaches to problems brought on by heterogeneity and the semi-automated preprocessing steps in the mining process. Therefore, in order to close research gaps within the data mining community, researchers and data scientists need to address future challenges in order to fully exploit the potential of Semantic Web mining.

References

[1] O. Menemencioglu and I. M. Orak, “A review on semantic web and recent trends in its applications,” in *2014 IEEE International Conference on Semantic Computing*, 2014, pp. 297–303.

[2] M. Keyvanpour, H. Hassanzadeh, and B. M. Khoshroo, “Comparative classification of semantic web challenges and data mining techniques,” in *2009 International Conference on Web Information Systems and Mining*, 2009, pp. 200–203.

[3] A. Aksac, O. Ozturk, and E. Dogdu, “A novel semantic web browser for user centric information retrieval: Person,” *Expert Systems with Applications*, vol. 39, no. 15, pp. 12 001–12 013, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417412005799

[4] A. Patel and S. Jain, “Present and future of semantic web technologies: a research statement,” *International Journal of Computers and Applications*, vol. 43, no. 5, pp. 413–422, 2021. [Online]. Available: https://doi.org/10.1080/1206212X.2019.1570666

[5] Q. K. Quboa and M. Saraee, “A state-of-the-art survey on semantic web mining,” *Intell. Inf. Manag.*, vol. 05, no. 01, pp. 10–17, 2013.

[6] B. Berendt, A. Hotho, and G. Stumme, “Towards semantic web mining,” in *The Semantic Web — ISWC 2002*, I. Horrocks and J. Hendler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 264–278.

[7] S. S. Dhenakaran and S. Yasodha, “Semantic web mining-a critical review,” in *Semantic Web Mining-A Critical Review*, 2011.

[8] K. Sridevi and R. Umarani, “A survey of semantic based solutions to web mining - ijettcs ...” 2012. [Online]. Available: https://www.yumpu.com/en/document/view/8325887/a-survey-of-semantic-based-solutions-to-web-mining-ijettcs-

[9] A. C. Reyes, Tarlac State University, Tarlac City, Philippines, and https://orcid.org/0000-0003-0458-9409, “The implementation of indigenous peoples education (IPED) program: Towards a proposed plan of action,” *APJAET - Journal ay Asia Pacific Journal of Advanced Education and Technology*, pp. 13–24, Sep. 2022.

[10] “Systematic reviews: Methods,” 2018. [Online]. Available: https://guides.lib.umich.edu/c.php?g=283340&p=1887811#s-lg-box-6206921

[11] B. Kitchenham, "Procedures for performing systematic reviews," *Keele, UK, Keele Univ.*, vol. 33, 08 2004.

[12] "Google scholar," 2022. [Online]. Available: https://scholar.google.com/

[13] 2022. [Online]. Available: https://www.acm.org/

[14] 2022. [Online]. Available: https://ieeexplore.ieee.org/Xplore/home.jsp

[15] J. Dudovskiy, "Literature review: Measures for validity - research methodology," Aug 2012. [Online]. Available: https://research-methodology.net/literature-review-measures-for-validity/#:~:text=According%20to%20Brown%20(2006)%20there,whole%20process%20of%20literature%20review

[16] B. A. Hassan, "Towards semantic web: Challenges and needs," 2021. [Online]. Available: https://arxiv.org/abs/2105.02708

[17] P. Hitzler, "A review of the semantic web field," *Commun. ACM*, vol. 64, no. 2, p. 76–83, jan 2021. [Online]. Available: https://doi-org.proxy.lnu.se/10.1145/3397512

[18] U. Yadav, Research Scholar, Department of Computer Engineering, YMCA University of Science and Technology, Faridabad, India, G. Singh Narula, N. Duhan, and V. Jain, "Ontology engineering and development aspects: A survey," *Int. j. educ. manag. eng.*, vol. 6, no. 3, pp. 9–19, May 2016.

[19] I. Osman, S. Ben Yahia, and G. Diallo, "Ontology integration: Approaches and challenging issues," *Information Fusion*, vol. 71, pp. 38–63, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253521000130

[20] C. Reyes-Peña and T. Mireya, "Ontology: Components and evaluation, a review," *Research in Computing Science*, vol. 148, pp. 257–265, 12 2019.

[21] H. Gorskis and Y. Chizhov, "Ontology building using data mining techniques," *Information Technology and Management Science*, vol. 15, 12 2012.

[22] J. A. Khan and S. Kumar, "Deep analysis for development of rdf, rdfs and owl ontologies with protege," in *Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization*, 2014, pp. 1–6.

[23] A. Jain, P. I. Khan, and D. B. Verma, "Secure and intelligent decision making in semantic web mining."

[24] M. Barati, Q. Bai, and Q. Liu, "Mining semantic association rules from rdf data," *Knowledge-Based Systems*, vol. 133, pp. 183–196, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705117303258

[25] A. C. Reyes, Tarlac State University, Tarlac City, Philippines, and https://orcid.org/0000-0003-0458-9409, "The implementation of indigenous peoples education (IPED) program: Towards a proposed plan of action," *APJAET - Journal ay Asia Pacific Journal of Advanced Education and Technology*, pp. 13–24, Sep. 2022.

[26] V. Nebot and R. Berlanga, "Finding association rules in semantic web data," *Knowledge-Based Systems*, vol. 25, no. 1, pp. 51–62, 2012, special Issue on New Trends in Data Mining. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705111001031

[27] G. Kesavaraj and S. Sukumaran, "A study on classification techniques in data mining," in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 2013, pp. 1–7.

[28] Garima, H. Gulati, and P. Singh, "Clustering techniques in data mining: A comparison," in *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2015, pp. 410–415.

[29] K. Singh, A. Kumar, and A. K. Yadav, "Semantic web mining: Issues and challenges," in *SEMANTIC WEB MINING: ISSUES AND CHALLENGES*, 2016.

[30] C. E. Dinucă and D. Ciobanu, "Web content mining," *Scholarly Journal*, vol. 12, p. 85–92, 2012. [Online]. Available: https://www-proquest-com.proxy.lnu.se/docview/2109278986/fulltextPDF/31AEn.d.A481F45EBPQ/1?accountid=14827

[31] A. Kumar and R. Singh, "A study on web structure mining," in *A Study on Web Structure Mining*, 2017.

[32] C. R. Varnagar, N. N. Madhak, T. M. Kodinariya, and J. N. Rathod, "Web usage mining: A review on process, methods and techniques," in *2013 International Conference on Information Communication and Embedded Systems (ICICES)*, 2013, pp. 40–46.

[33] D. Jeon and W. Kim, "Development of semantic decision tree," in *The 3rd International Conference on Data Mining and Intelligent Information Technology Applications*, 2011, pp. 28–34.

[34] Z. Abedjan and F. Naumann, "Context and target configurations for mining rdf data," in *Proceedings of the 1st International Workshop on Search and Mining Entity-Relationship Data*, ser. SMER '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 23–24. [Online]. Available: https://doi-org.proxy.lnu.se/10.1145/2064988.2064998

[35] P. Shvaiko and J. Euzenat, "Ontology matching: State of the art and future challenges," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 1, pp. 158–176, 2013.

[36] P. Shvaiko, *Ontology Matching*. New York, NY: Springer New York, 2014, pp. 1209–1211. [Online]. Available: https://doi.org/10.1007/978-1-4614-6170-8_123

[37] J. D. Velásquez, L. E. Dujovne, and G. L'Huillier, "Extracting significant website key objects: A semantic web mining approach," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 8, pp. 1532–1541, 2011, semantic-based Information and Engineering Systems. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095219761100025X

[38] A. C. Reyes, Tarlac State University, Tarlac City, Philippines, and https://orcid.org/0000-0003-0458-9409, "The implementation of indigenous peoples education (IPED) program: Towards a proposed plan of action," *APJAET - Journal ay Asia Pacific Journal of Advanced Education and Technology*, pp. 13–24, Sep. 2022.

[39] M. Barati, Q. Bai, and Q. Liu, "Swarm: An approach for mining semantic association rules from semantic web data," in *PRICAI 2016: Trends in Artificial Intelligence*, R. Booth and M.-L. Zhang, Eds. Cham: Springer International Publishing, 2016, pp. 30–43.

[40] M. Allahyari, K. J. Kochut, and M. Janik, "Ontology-based text classification into dynamically defined topics," in *2014 IEEE International Conference on Semantic Computing*, 2014, pp. 273–278.

[41] "Hyperlink induced topic search (hits) algorithm using networxx module: Python," Feb 2021. [Online]. Available: https://www.geeksforgeeks.org/hyperlink-induced-topic-search-hits-algorithm-using-networxx-module-python/

[42] C. M. Keet, A. Ławrynowicz, C. d'Amato, A. Kalousis, P. Nguyen, R. Palma, R. Stevens, and M. Hilario, "The data mining optimization ontology," *Journal of Web Semantics*, vol. 32, pp. 43–53, 2015.

[Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570826815000025

[43] "Extensible markup language (xml)." [Online]. Available: https://www.w3.org/XML/

[44] "Owl 2 web ontology language structural specification and functional-style syntax (second edition)." [Online]. Available: https://www.w3.org/TR/owl2-syntax/

[45] D. Dou, H. Wang, and H. Liu, "Semantic data mining: A survey of ontology-based approaches," in *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)*, 2015, pp. 244–251.

A Appendix

The appendix holds tables with keywords and definitions as well a a reference list for primary studies in the SLR.

Keyword	Definition	Ref.
URI	Responsible for identifying and encoding resources and its identification.	[5]
XML, NS, XML schema	Separates data content, structure, and performance format based on linguistics and represents them in a standard format language.	[43]
RDF and RDF schema	Provides a common language for storing resources on the web using their own domain vocabularies.	[5]
Ontology vocabulary	Reveals semantics among information by defining the knowledge shared and the semantic relations within different types of information.	[44]
Logic	Row 5, Handles logical reasoning by providing axioms and inference principles.	[5]
Proof, Trust	Provides security to the web by using encryption and digital signature mechanisms to identify changes in documents.	[5]

Table A: Semantic Web Keywords

#	Title	Author	Year	Ref.
S1	Ontology Engineering and Development Aspects: A Survey	Yadav, Narula, Duhan, Jain	2016	[21]
S2	Ontology Integration: Approaches and Challenging Issues	Osman, Yahia, Diallo	2021	[22]
S3	Ontology Building Using Data Mining Techniques	Gorskis, Chizhov	2012	[24]
S4	Deep analysis for development of RDF, RDFS and OWL ontologies with protege	Khan, Kumar	2014	[25]
S5	Secure and Intelligent Decision Making in Semantic web mining	Ankita, Ilyas, Verma Bhupendra	2011	[26]
S6	Mining Semantic association rules from RDF data	Barati, Bai, Liu	2017	[27]
S7	Improving RDF Data Through Association Rule Mining	Abedjan, Naumann	2013	[28]
S8	A Study on Classification Techniques in Data Mining	Kesavaraj, Sukumaran	2013	[29]
S9	Clustering Techniques in Data Mining: A Comparison	Garima, Gulati, Singh	2015	[30]
S10	Semantic Web Mining: Issues and Challenges	Singh, Kumar, Yadav	2016	[31]
S11	Web Content Mining	Dinuca, Ciobanu	2012	[32]
S12	A study on Web Structure Mining	Kumar, Singh	2017	[33]
S13	Finding Association rules in semantic web data	Nebot, Berlanga	2012	[35]
S14	Development of semantic decision tree	Jeon, Kim	2011	[36]
S15	Context and target configurations for mining RDF data	Abedjan, Naumann	2011	[37]
S16	Ontology Matching: State of the Art and Future Challenges	Shvaiko, Euzenat	2013	[38]
S17	Ontology Matching	Shvaiko	2014	[39]
S18	Extracting significant Website Key Objects: A Semantic Web mining approach	Velasquez, Dujovne, L'Huillier	2011	[40]
S19	On ontology-driven document clustering using core semantic features	Fodeh, Punch, Tan	2011	[41]
S20	SWARM: An approach for Mining Semantic Association Rules from Semantic Web Data	Barati, Bai, Liu	2016	[42]
S21	Ontology-based Text Classification into Dynamically Defined Topics	Allahyari, Kochut, Janik	2014	[43]
S22	The Data Mining Optimization Ontology	Keet, Lawrynowicz, d'Amato, Kalousis, Nguyen, Palma, Stevens, Hilario	2015	[45]
S23	Using Ontologies in Semantic Data Mining with SEGS and g-SEGS	Lavrac, Vavetic, Soldatova, Trajkovski, Novak	2011	[12]
S24	Semantic data mining: A survey of ontology-based approaches	Dou, Wang, Lio	2015	[[45]]
S25	A survey of Semantic based Solutions to Web Mining	Sridevi, Umarani	2013	[10]
S26	A State-of-the-Art Survey on Semantic Web Mining	Quboa, Saraee	2012	[5]
S27	Present and future of semantic web technologies: a research statement	Patel	2018	[4]
S28	A novel semantic web browser for user centric information retrieval: PERSON	Aksac, Ozturk, Dogdu	2012	[3]