ISSN: 1803-3814 (Printed), 2571-3701 (Online)
https://doi.org/10.13164/mendel.2023.2.111

Automated Semantic Annotation Deploying Machine Learning Approaches: A Systematic Review

Wee Chea Chang[1,✉], **Anbuselvan Sangodiah**[2]

[1]Faculty of Technology and Applied Science, Open University Malaysia, Malaysia
[2]School of Computing, Faculty of Computing and Engineering, Quest International University, Malaysia
weechea@gmail.com✉, anbuselvan.sangodiah@qiu.edu.my

Abstract

Semantic Web is the vision to make Internet data machine-readable to achieve information retrieval with higher granularity and personalisation. Semantic annotation is the process that binds machine-understandable descriptions into Web resources such as text and images. Hence, the success of Semantic Web depends on the wide availability of semantically annotated Web resources. However, there remains a huge amount of unannotated Web resources due to the limited annotation capability available. In order to address this, machine learning approaches have been used to improve the automation process. This Systematic Review aims to summarise the existing state-of-the-art literature to answer five Research Questions focusing on machine learning driven semantic annotation automation. The analysis of 40 selected primary studies reveals that the use of unitary and combination of machine learning algorithms are both the current directions. Support Vector Machine (SVM) is the most-used algorithm, and supervised learning is the predominant machine learning type. Both semi-automated and fully automated annotation are almost nearly achieved. Meanwhile, text is the most annotated Web resource; and the availability of third-party annotation tools is in-line with this. While Precision, Recall, F-Measure and Accuracy are the most deployed quality metrics, not all the studies measured the quality of the annotated results. In the future, standardising quality measures is the direction for research.

Keywords: *Semantic Web, Semantic Annotation Automation, Machine Learning, Quality Metrics, Systematic Review.*

Received: 01 August 2023
Accepted: 15 September 2023
Online: 18 September 2023
Published: 20 December 2023

1 Introduction

Web 2.0 is the current standard of the World Wide Web (i.e., the Web), in which the basics of collaborative content creation were laid out. Since year 2004, Web 2.0 has resulted in information explosion. This is due to the growing amount of mobile internet access, authoring tools such as Wikis, and social networking platforms such as Facebook. According to International Data Corporation (IDC), there was 79 zettabytes of data created in year 2021. Consequently, this has increased its disorganisation and complexity, leading to huge amount of untapped information (80% - 90% of the total amount of data generated) and imprecise query outcomes [22], [48].

Web 3.0 is the third generation of the Web's evolution, aiming at addressing these shortcomings. Semantic Web (an important building block of Web 3.0) is the mission to make the Web resources machine-readable (and thus also link-able and relatable among all the Web resources), thereby creating a "Web-of-Data". Semantic annotation binds machine-understandable formal descriptions (ontologies) into the Web resources such as text, images and Web services. Thus, the success of Semantic Web requires wide availability of semantically annotated Web resources. However, due to the limited annotation scope and capability available, there is still a huge amount of unannotated Web resources. Automated semantic annotation provides the help in reducing human intervention throughout the process, hence achieving the desired annotation speed, scalability and consistency, while reducing human mistakes. For this, machine learning approaches are the more focused field of studies since human factors are kept to the minimum, and there are a lot of machine learning algorithms, studies and applications that can be leveraged on.

Since the scope of semantic annotation automation is wide, there is a need to have a centralised, objective and comprehensive survey that covers this topic. However, such a survey is either still missing or is outdated (i.e., published more than 10 years ago). This survey systematically reviews the existing literature for the state-of-the-art of semantic annotation automation driven by machine learning approaches. It covers topics on the degrees of automation, the type of machine learning approaches, the algorithms, the application domains, the available third- party tools and the quality indicators.

The remainder of this paper is organised as follows: Section 2 describes the main knowledge of semantic annotation automation; Section 3 details about the research method (i.e., Systematic Review) and steps involved based on the Research Questions; Section 4 analyses the extracted data; Section 5 discusses the analysed results, the threats to validity, the future works and the final conclusion.

2 Related Work

2.1 The Semantic Web and Semantic Annotation

Tim Berners-Lee described the Semantic Web as "an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in corporations" [11], [72]. It was the idea of having information on the Web defined and linked in a way that it can be used for more effective discovery, automation, integration, and reuse across various applications [27], [40].

Semantic Web is hence a Web that consists of human-readable parts and sections with formats accessible by machines for automated processing. It is based on two fundamental concepts: *ontology* and *annotation* [40], [69]. An ontology is "a specifi- cation of a representational vocabulary for a shared domain of disclosure, including definitions of classes, relations, functions and other objects" [9], [26]. In order for machines to understand semantic meanings, those meanings and relationships have to be established through common standards of Resource Description Framework (RDF) and Web Ontology Language (OWL). Fig. 1 depicts the "Semantic Web Layer Cake" that illustrates the architecture of the Semantic Web; while Fig. 2 simplifies the concept of Semantic Web.

Annotation is the process of allocating some labels to the original data for data interpretation and automatic description [1]. Semantic annotation (also known as Semantic Web annotation) is thus the process in which some necessary information (in the forms of RDF and OWL) is added to Web resources (such as text and images) to reflect the relationship between ontology class concepts and the Web resources. Such annotation enables target information to be easily searched and classified by the machine.

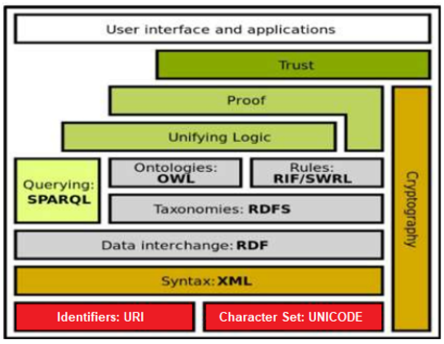

Figure 1: Semantic Web Layer Cake [79].

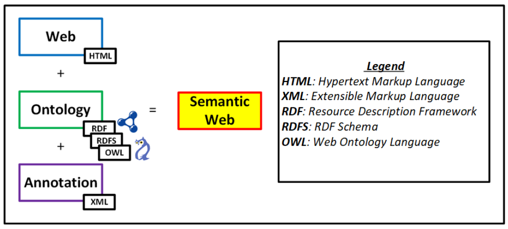

Figure 2: Simplified View of Semantic Web.

The World Wide Web Consortium (W3C) defines the main semantic annota- tion standards: Resource Description Framework (RDF), RDF-Schema (RDFS) and Web Ontology Language (OWL). RDF is used to make statement about instances through the form of triple (subject, predicate, object); RDFS defines schema and subclass hierarchies; and OWL is the ontology language used to formulate additional background knowledge [40], [66]. These standards are slotted into the Semantic Web Layer Cake (Fig. 1). Fig. 3 is an example on how RDF is realised in Extensible Markup Language (XML).

2.2 Degrees of Automation

The degree of automation defines the automaticity of semantic annotation, and it can be measured as manual, semi-automatic and fully automatic [14], [40]. Manual annotation (Fig. 4) is the process of reading an input Web resource and extracting new information with human participation. This is a type of formal annotation with human-computer interaction. Manual annotation can be conveniently done today with authoring tools such as Semantic Word [10], [65]. Manual annotation is more precise compared to automated annotation. However, it is very labour-intensive, requiring an annotator to be a domain expert, time-consuming and often full of errors. Manual annotation is too expensive to achieve the economy-of-scale. Hence, it is only suitable for small-scale annotation, or in cases where semantic annotation is done in parallel with the development of a new Web resource.

Semi-automatic annotation process (Fig. 5) needs human intervention at some annotation level. Example tools include GATE and Semantator [63], [85]. Most of the semi-automatic annotation systems are derived from supervised machine learn- ing algorithms that involve extensive training, hence the human intervention. Semi- automatic annotation is fast and robust in finding the semantic relationship between the annotating data and the targeted annotated document. Human enrolment pro- vides a significant advantage of adopting new features and new domains.

Fully automatic semantic annotation (Fig. 6) is a high-level semantic annota- tion. Fully automatic systems are highly trained for its automaticity [59]. To train this type of system, a large amount of labelled data and rule sets are required when deploying supervised machine learning algorithms, and this is an expensive process. To minimise these issues, unsupervised systems have tried methodologies and exper-

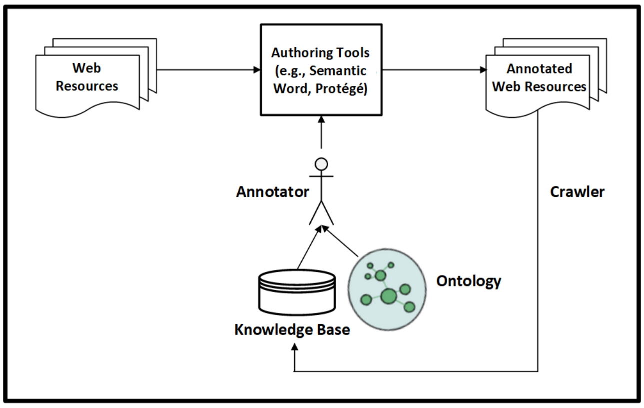

Figure 4: Manual Semantic Annotation [59].

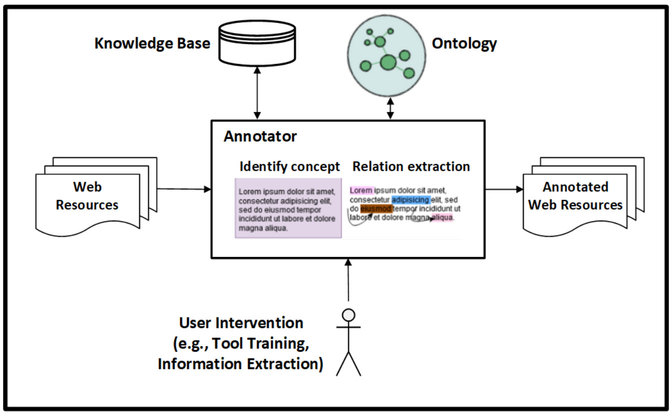

Figure 5: Semi-automatic Semantic Annotation [59].

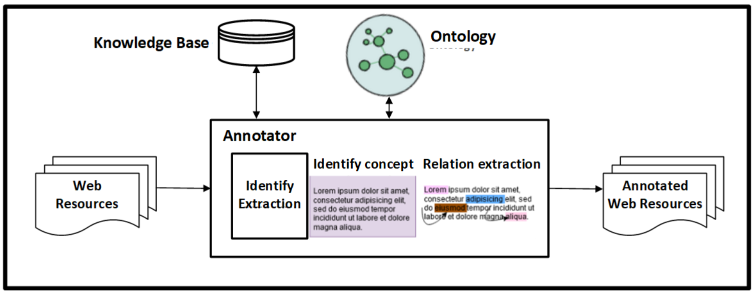

Figure 6: Fully Automatic Semantic Annotation [59].

Table 1: Advantages and Disadvantages on Annotation Techniques [62].

Annotation Techniques	**Manual**	**Automatic**	**Semi-automatic**
Advantages	The most accurate annotation	The most efficient, the least time	Quality of the annotation improves in an interactive manner after human correction
Disadvantages	Time consuming (expensive), difficult, subjective, inconsistent	Error-prone, the less accurate annotation	Less time than automatic annotation, greater time than manual annotation

Table 2: Contingency Matrix for the Annotation Process [78].

<table>
<tr><td></td><td></td><td colspan="2">Current Annotation</td><td></td><td></td></tr>
<tr><td></td><td></td><td>Positive Annotation</td><td>Negative Annotation</td><td></td><td rowspan="4">F1-score = (2*P*R) / (P+R)</td></tr>
<tr><td rowspan="2">Predicted Annotation</td><td>Positive Prediction</td><td>TP
(True Positive)
"hit"</td><td>FP
(False Positive)
"false alarm, overestimation"</td><td>Precision (P) = TP / (TP+FP)</td></tr>
<tr><td>Negative Prediction</td><td>FN
(False Negative)
"miss, underestimation"</td><td>TN
(True Negative)
"correct rejection"</td><td></td></tr>
<tr><td></td><td></td><td>Recall (R) = TP / (TP+FN)</td><td>Accuracy = (TP+TN) / Total</td><td></td></tr>
</table>

- Comprehensive sources with explicit search approach.
- Selection and evaluation of literature are criterion-based with evaluation scoring.

The Systematic Review protocol guidelines and templates are based on the works of [12], [34], [35]. Fig. 11 summarises the Systematic Review protocol and the actual sub-tasks to be carried out in each of the steps.

3.1 Research Questions

The goal of this research is to support the following primary Research Question:

"How far has the automated semantic annotation been achieved through machine learning?"

Based on this primary question, specific Research Questions are identified (Table 3).

3.2 Source and Study Selection

As a necessary starting point, Systematic Review aims to find all primary studies related to the Research Questions identified. The selection criterions of sources include:

- Trusted source
- Availability of text in English
- Availability of contemporary collection of papers (i.e., from 2013 to 2022)
- Advanced search capabilities (filtered by title, abstract, keywords)
- Abundance of publications medium (e.g., journals, conferences, workshops, etc.)
- Quality of the querying engine of the source

Based on these requirements, the following electronic databases are selected:

1. Google Scholar (http://scholar.google.com)
2. ACM Digital Library (http://dl.acm.org)
3. ProQuest Dissertations & Theses Global (https://www.proquest.com)
4. IEEEXplore Digital Library (http://ieeexplore.ieee.org)

Based on the identified Research Questions and the Systematic Review guidelines, a search string is defined (Fig. 12). In case similar or duplicate studies are detected, the latest publications are selected. Based on the guidelines of [34], inclusion criteria and exclusion criteria defined for this Systematic Review are shown in Table 4.

Fig. 13 details all the search stages involved in this Systematic Review. In the first stage of the process, the search string is executed for each of the four electronics databases. It is necessary to adjust the search string according to the unique requirements for each of the electronics databases. The search was conducted in September 2022. 1492 primary studies have been selected.

In order to limit the studies to the most recent ones, only studies published after January 2013 (IC2 of Table 4) are considered. Besides, only studies written in English are considered (EC3 of Table 4). This has resulted in a reduction into 501 studies. Next, only primary studies are considered (IC1, EC1 of Table 4) and thus 16 secondary studies have been eliminated to a total of 485 studies. Subsequently, 70 duplicated studies are removed, resulting in 415 studies brought forward to the next step. 174 studies without full text availability (EC4 of Table 4) are discarded next.

The next stages involve iterative application of the inclusion and exclusion criteria (Table 4). Firstly, title, keywords and areas of knowledge are gone through, and those that are either not meeting inclusion criteria or are not relating to the Research Questions are excluded. After discarding 130 studies, 111 primary studies have remained.

Next, the abstracts of the studies are read. The criteria for inclusion and exclusion (Table 4) are applied again. This results in the exclusion of another 45 studies. The next iteration involves reading the introduction and conclusion sections of the remaining 66 studies, which further filters out 16 studies.

By considering the list of references of the secondary

```
(
 ("SEMANTIC ANNOTATION" OR
  "SEMANTIC WEB")
                        AND
 ("AUTO"                OR
  "MACHINE LEARNING"    OR
  "DEEP LEARNING"       OR
  "SUPERVISED"          OR
  "UNSUPERVISED"        OR
  "TOOL"                OR
  "QUALITY")
)
```

Figure 12: Search String.

utilisation of the 8 possible scores. Table 6 lists the final selected studies, together with the results of the quality evaluation incurred.

3.4 Data Extraction

The data extraction phase involves collecting information relevant to the Research Questions from the selected studies. Table 7 shows the Data Extraction Form created based on the guidelines of [35] and the complete reading of the final 40 selected studies.

The general information to be extracted are: year, country, and publications medium. By taking Research Questions into account, the specific data includes algorithms, degrees of automation, domains targeted and the areas of application, commercial or open-source tools, and quality indicators.

4 Result Analysis

4.1 Year of Publication

The years of publication in this review are constrained between the years of 2013 and 2022 (IC2 of Table 4). Most of the studies were published in 2016 (17.5%) and 2014 (17.5%), followed by 2017 (12.5%), 2019 (12.5%), 2015 and 2020 (10.0% re- spectively), 2018 (7.5%), 2013 and 2021 (5.0% respectively) and finally 2022 (2.5%). The number of publications over the years is depicted in the line chart in Fig. 14.

4.2 Country

As depicted in the bar chart of Fig. 15, China (20.0%) has contributed the most of the selected primary studies. By considering the continents, Europe has the most studies selected (45.0%), while Africa has 0 studies selected (0.0%). To complete the list, Asia represents 35.0% of the selected studies, followed by North America and Oceania (both 7.5% respectively) and South America (5.0%).

4.3 Publications Medium

The types of publications medium adopted in this review are journals, conferences, workshops, book chapters and electronics archives (Item 4 of Table 7). Book chapters here include master's theses and PhD's dissertations. The electronics archives are for studies stored in the Research Square platform (researchsquare.com) that are not published elsewhere. According to the pie chart of Fig. 16, most of the studies were published in journals (47%), followed by conferences (32%), book chapters (15%), and workshops and electronic archives (both % respectively).

4.4 RQ1: Algorithms

The goal of this Research Question is to identify the types of machine learning and the machine learning algorithms deployed in the process of semantic annotation automation. Three situations are observed: the first refers to the use of a unitary or single algorithm, while the second is the use of a combination of algorithms. The final situation refers to the case in which the actual algorithm deployed is not disclosed.

There are 2 studies (i.e., [18] and [53]) that did not disclose the actual algorithms deployed. These 2 studies instead focused on the overall flow optimization of the semantic annotation.

For the remaining 38 studies, 20 studies (52.63%) deployed a single algorithm as the basis for automated semantic annotation. Table 8 shows the distribution of algorithms. There has been no predominant algorithm used, though both Random Forest and K-Means were deployed in more than one studies (i.e., 2 studies respectively). However, supervised learning is the predominant type of machine learning as there are 14 studies altogether (i.e., 70.00% out of 20 studies here). On the other hand, unsupervised learning and semi-supervised learning types are only deployed by 5 studies and 1 study respectively.

For the remaining 18 studies that combined machine learning algorithms, it has been observed that SVM was the more preferred algorithm, followed by Bag-of-Word (BOW) and Latent Semantic Analysis (LSA). Again, supervised learning is the predominant type that partially contributes to 150.00% of the overall distribution. Both unsupervised and semi-supervised machine learning , meanwhile, partially contribute 77.78% and 22.22%, as detailed in Table 9.

By combining these 18 studies with combined algorithms to the 20 studies with unitary algorithms (Table 10), the top three algorithms are SVM (9.09%), BOW (6.06%) and K-Means (6.06%). The predominant use of SVM (almost one-tenth overall) demonstrates that the processing cost has become an increasingly small hindrance to the feasibility of research works. SVM is known for its robustness and optimal accuracy values, and it is relatively agnostic to outliers. Furthermore, the memory-efficient nature of SVM is well-accepted when dealing with large datasets. BOW and K-Means, meanwhile, are simpler and easier to use. BOW is usually used by researchers to create the first prototype model for textual Web resource, while K-Means is deployed in document clustering and image segmentation. K-Means also benefits from its

Table 6: Studies Selected through Quality Evaluation.

ID	References	QC1	QC2	QC3	QC4	QC5	QC6	QC7	QC8	Total	%
S001	[58]	0.5	0.5	1.0	1.0	1.0	1.0	1.0	1.0	7.0	88%
S002	[50]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	7.5	94%
S003	[83]	1.0	1.0	1.0	1.0	0.5	1.0	1.0	0.5	7.0	88%
S004	[18]	0.5	1.0	1.0	0.0	1.0	1.0	1.0	1.0	6.5	81%
S005	[24]	0.0	0.5	1.0	1.0	1.0	0.5	1.0	0.5	5.5	69%
S006	[84]	1.0	0.5	1.0	0.0	1.0	1.0	1.0	0.0	5.5	69%
S007	[37]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S008	[41]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S009	[31]	1.0	0.5	1.0	0.0	1.0	1.0	1.0	1.0	6.5	81%
S010	[89]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	7.0	88%
S011	[86]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S012	[19]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S013	[15]	1.0	1.0	1.0	0.0	1.0	1.0	0.5	0.5	6.0	75%
S014	[75]	1.0	0.5	1.0	1.0	1.0	1.0	1.0	0.5	7.0	88%
S015	[16]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	7.5	94%
S016	[42]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S017	[44]	1.0	0.5	1.0	0.5	1.0	1.0	1.0	0.0	6.0	75%
S018	[3]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S019	[76]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S020	[13]	1.0	0.5	1.0	0.5	1.0	1.0	0.5	0.0	5.5	69%
S021	[64]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S022	[30]	1.0	0.5	1.0	0.5	1.0	1.0	1.0	0.5	6.5	81%
S023	[47]	0.5	0.5	1.0	1.0	1.0	1.0	1.0	0.5	6.5	81%
S024	[88]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S025	[82]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S026	[74]	1.0	0.5	1.0	1.0	1.0	1.0	1.0	0.5	7.0	88%
S027	[55]	1.0	0.5	1.0	0.5	0.0	0.0	0.5	1.0	4.5	56%
S028	[21]	0.5	0.5	0.5	1.0	0.5	0.5	1.0	1.0	5.5	69%
S029	[8]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S030	[32]	0.5	0.5	1.0	0.0	1.0	1.0	1.0	1.0	6.0	75%
S031	[7]	1.0	1.0	1.0	1.0	0.5	1.0	0.5	1.0	7.0	88%
S032	[29]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S033	[43]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S034	[81]	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	7.0	88%
S035	[5]	1.0	0.5	1.0	1.0	1.0	1.0	1.0	0.5	7.0	88%
S036	[28]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S037	[6]	1.0	0.5	0.5	0.5	0.5	1.0	1.0	1.0	6.0	75%
S038	[53]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S039	[49]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%
S040	[73]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	8.0	100%

Table 7: Data Extraction Form.

	Information for Extraction	Description	Relevance
1.	ID	Unique study identifier	General
2.	Year	Year of publication	General
3.	Country	Country of origin	General
4.	Publications Medium	Journal, conference, book chapter, workshop, electronics archive	General
5.	Algorithms	Type of machine learning, its category of application, and the exact algorithm used	RQ1
6.	Degrees of Automation	Semi-automated vs. fully automated	RQ2
7.	Domains Targeted and Areas of Application	Domain targeted for the semantic annotation and the areas of application	RQ3
8.	Commercial or Open-source Tools	Automated semantic annotation tools	RQ4
9.	Quality Indicators	Quality consideration. Indicators to quality analysis such as accuracy, precision, recall and F1-score	RQ5

robustness to outliers and efficiency in handling large datasets.

Supervised machine learning algorithms are predominantly deployed, spanning over 60% of the overall distribution. This reveals that the supervised machine learning type is still the mainstream of the research direction since accuracy is critical in the field of semantic annotation automation.

Table 10: Overall Distributions of Machine Learning Algorithms.

Type of ML	Algorithm	Studies	Quantity	% (General-Algorithm)	Quantity (ML Type)	% (General-ML Type)
Supervised	SVM	[31], [89] [76], [13], [21], [43]	6	9.09	42	63.64
	Random Forest	[8], [29], [43]	3	4.55		
	Gabor Filtering	[37], [21]	2	3.03		
	Naive Bayes	[37], [44]	2	3.03		
	KNN	[83], [37]	2	3.03		
	Decision Trees	[32], [43]	2	3.03		
	Inception-V3	[58]	1	1.52		
	WED	[37]	1	1.52		
	Mask-RCNN	[58]	1	1.52		
	MKL	[31]	1	1.52		
	Mean shift algorithm	[83]	1	1.52		
	EM	[89]	1	1.52		
	RLSD	[86]	1	1.52		
	RNN	[86]	1	1.52		
	R-CNN	[86]	1	1.52		
	Adaboost	[15]	1	1.52		
	BPNN	[15]	1	1.52		
	Genetic Algorithm	[16]	1	1.52		
	Levenshtein Distance	[64]	1	1.52		
	Common Words	[64]	1	1.52		
	g-TKSE	[30]	1	1.52		
	LSTM	[47]	1	1.52		
	Time Series	[82]	1	1.52		
	OASA	[88]	1	1.52		
	Maximum entropy	[55]	1	1.52		
	Ontology-based	[24]	1	1.52		
	MSER	[37]	1	1.52		
	Shallow NLP	[7]	1	1.52		
	MNL	[43]	1	1.52		
	DeepWalk	[6]	1	1.52		
	Logistic regression	[49]	1	1.52		
Unsupervised	BOW	[50], [74], [5], [6]	4	6.06	19	28.79
	K-Means	[84], [31], [42], [28]	4	6.06		
	LSA	[89], [74], [28]	3	4.55		
	KDS	[41]	1	1.52		
	FFCA	[19]	1	1.52		
	FRCA	[19]	1	1.52		
	Generic Programming	[75]	1	1.52		
	Microservices-based	[3]	1	1.52		
	DBN	[76]	1	1.52		
	Apriori	[41]	1	1.52		
	LDA	[28]	1	1.52		
Semi-supervised	Skip-gram	[5], [6]	2	3.03	5	7.58
	SSAE	[81]	1	1.52		
	NLP with Similarity Measure	[73]	1	1.52		
	Context-based Graph Filtering	[73]	1	1.52		
Total			**66**	100.00	66	100.00

the more sophisticated (albeit more complicated) fully automated annotation has been gaining a significant footprint for the past 10 years.

4.6 RQ3: Domains Targeted and the Area of Application

The purpose of this Research Question is to identify the distribution of the domains targeted by the selected studies. These domains are text, images and Web services. Table 12 shows the distribution of the domains targeted, together with the areas of application involved for each of the domains. Automated se-

Table 11: Distribution of the Degrees of Automation.

Degree of Automation	Studies	Quantity	%
Semi-automated	[58], [50], [83], [18], [84], [31], [89], [86], [15], [16], [42], [44], [64], [47], [82], [74], [55], [21], [8], [7], [43], [28], [53]	23	57
Fully automated	[24], [37], [41], [19], [75], [3], [76], [13], [30], [88], [32], [29], [81], [5], [6], [49], [73]	17	43
Total		40	100

Table 12: Distribution of Domains Targeted and Areas of Application.

Domains	Studies	Quantity	%	Areas of Application
Text	[18], [24], [84], [19], [16], [44], [3], [13], [47], [47], [74], [55], [8], [7], [29], [43], [5], [28], [6], [53], [49], [73]	22	55.0	linguistic, sensor data, activity recognition, educational content, academic papers, online training, unstructured text document, configuration fine tuning, food, health & nutrition information, IoT data, Web documents, cell phone location, document categorisation, legal & encyclopedia data
Images	[58], [83], [37], [31], [89], [86], [15], [76], [21], [32], [81]	11	27.5	SAR images, medical images, compound object images, news images, historical catalogues, image retrieval, tourism images, generic images, emotional annotation on scene images, satellite images
Web Services	[50], [41], [75], [42], [64], [30], [88]	7	17.5	WSDL, SOA, OWL-S, SSN, SAWSDL, IoT, quality-aware Web service composition
Total		40	100.0	

tributed to the following reasons:

- Maturity level of the other types of machine learning is not on par yet.
- Accurate annotation is of utmost importance in the field of semantic annotation, specifically in the areas such as medical, security and engineering.

Having said that, the deployment of unsupervised and semi-supervised learnings is gaining meaningful acceptance, contributing to more than 35% of the selected studies. This is in line with [59] which stated that the research has started moving towards unsupervised or semi-supervised machine learning methods due to the cost-prohibitive and laborious nature of supervised learning.

In answering RQ2, there are equal contributions between semi-automated annotation and fully automated annotation identified. The semi-automated annotation has an advantage of producing high quality results. It is also more adaptive to new changes due to human intervention. Hence, this is best used in exploring new types of Web resources as well as new areas of application. On the other hand, fully automated annotation has been gaining a significant focus for the past 10 years, with almost half of the selected studies achieving full automation. This is important as fully automated annotation is the one capable of annotating the huge amount of unannotated Web resources.

For RQ3, textual resources are the most semantically annotated Web resource. This is expected as text is the most widely available Web resource. There have been a lot of researches and studies carried out on text processing and natural language processing since the 1980s [17]. The finding of third-party tools (in answering RQ4) is in line with this, and therefore the supply-and-demand theory holds here.

When studying the available third-party tools (RQ4), this review identifies that open-source solutions are much more dominant over the commercial solutions. This reflects the current trend in this field, in which it is still actively researched and collaborative. It needs time for the perfection of such tools before commercial alternatives gradually step in. The availability of matured third-party tools is important to push the semantic annotation automation forward.

RQ5 emphasises on the importance of quality results. The awareness of quality continues to be an issue. Quality measure is still observed as "optional" to some of the studies (as much as 10% of the selected studies). This should not be taken lightly especially in this field of semantic annotation automation, as the end goal is to automatically annotate the huge amount of unannotated Web resources. This review identifies that Precision, Recall, F-Measure and Accuracy are the most used quality metrics, and these are in line with studies done by [61]. In order to perform tasks such as cross-tool quality and performance benchmarking, it is imperative to have standardised quality indicators so that all parties can speak on the same language with common calibration of expectations and understandings. Hence, the efforts in standardising the quality measures shall be taken seriously by the governing bodies such as W3C now.

5.2 Threats of Validity

Validity of the results is a main concern in studies deploying Systematic Review [68]. Here, threats to construct and internal validity [80] are discussed. Construct validity is about whether or not the implementation of a Systematic Review matches its initial objectives. The search string (Figure 12) is identified to be of main concern. This search string was derived from the Research Questions. However, the thoroughness of the keywords used is not guaranteed. Although well-established electronics databases were queried for the relevant primary studies, other sources queried with different keywords may still return relevant primary studies. However, this is not taken into consideration in this Systematic Review.

Internal validity is the extent to which the design and conduct of the study are likely to prevent systematic error [35]. The point of concern is on the data extraction. When extracting data from the selected studies,

Table 14: Distribution of Deployed Quality Metrics.

Measure	Studies	Quantity	% (General)
Precision	[50], [24], [41], [31], [89], [86], [19], [16], [44], [76], [13], [47], [81], [74], [55], [21], [8], [7], [29], [5], [6], [53], [49], [73]	24	60.0
Recall	[50], [24], [41], [86], [19], [16], [44], [13], [47], [88], [74], [55], [21], [8], [7], [29], [5], [6], [53], [49], [73]	21	52.5
F-Measure/ F-Score/ F1-Measure/ F1-Score	[50], [41], [19], [16], [44], [13], [47], [88], [74], [8], [7], [5], [53], [49], [73]	15	37.5
Accuracy	[58], [83], [15], [30], [8], [32], [43], [81], [28], [53], [49]	11	27.5
Not Applicable	[18], [42], [3], [64]	4	10.0
F1-AUC	[86], [16]	2	5.0
Mean Average Precision	[31], [86]	2	5.0
Correlation Degree on Nodes	[50]	1	2.5
Small-worldness	[50]	1	2.5
Recognition Rate	[84]	1	2.5
Weighted Error Rates	[37]	1	2.5
Overall Success Rate	[37]	1	2.5
Scale-Freeness	[50]	1	2.5
Matchmaking Quality	[75]	1	2.5
Response Time	[75]	1	2.5
Cost	[75]	1	2.5
Reliability	[75]	1	2.5
Availability	[75]	1	2.5
Certainty Evaluation	[47]	1	2.5
Inter-rater Reliability	[82]	1	2.5
Running Time	[81]	1	2.5
Pearson Correlation	[53]	1	2.5
Spearman Correlation	[53]	1	2.5

some level of self-interpretation was required whenever the data of interest was not clearly expressed. All steps followed in this Systematic Review were executed twice to minimise such error.

5.3 Directions for Future Research

For future research, the goal is to carry out a more thorough review that spans across wider facets of semantic annotation automation.

Firstly, more types of Web resources in the forms of videos and audios will be factored in, as the abundant availability of these Web resources (on plaforms such as YouTube and Spotify) should be seriously taken into account. Next, reinforcement learning will be another type of machine learning that will be taken into account in the future, as its application is getting more traction in the Natural Language Processing (NLP) area, in which it is one of the crucial parts of textual annotation [51].

Besides, studies on standardising the quality measures is an important direction for the future research in the field of semantic annotation automation, as the success of the Semantic Web highly depends on the high-quality and measurable annotated outcomes.

5.4 Conclusion

This review summarises the state-of-the-art in the field of semantic annotation au- tomation deploying machine learning approaches. It focuses on answering the iden- tified Research Questions (Section 3.1). 40 primary studies are selected through Systematic Review.

The use of unitary and combinations of algorithms is observed. Supervised learn- ing is the more predominant machine learning type, while SVM is the most preferred algorithm. Meanwhile, both semi-automated and fully automated annotation are almost equally achieved.

Text is the main Web resource to be semantically annotated, in-line with the availability of third-party tools. As for the third-party tools, the availability of open-source tools outweighs the commercial tools. Quality measurement is not performed by all of the selected studies, and this must be put to a stop in ensuring the success of Semantic Web. Meanwhile, Precision, Recall, F-Measure and Accuracy are the main quality metrics used.

References

[1] Oxford learner's dictionaries, 2022. https://www.oxfordlearnersdictionaries.com.

[2] Achimugu, P., Selamat, A., Ibrahim, R., and Mahrin, M. N. A systematic literature

[29] Gábor, K., Zargayouna, H., Buscaldi, D., Tellier, I., and Charnois, T. Semantic annotation of the acl anthology corpus for the automatic analysis of scientific literature. In *LREC* (2016), pp. 3694–3701.

[30] Hassani, A., Montori, F., Liao, K., Haghighi, P., Jayaraman, P., and Georgakopoulos, D. Informa: A tool for classification and semantic annotation of iot datastreams. In *2021 IEEE 7th World Forum on Internet of Things (WF-IoT)* (2021), pp. 223–228.

[31] Hou, A., Wang, C., Guo, J., Wu, L., and Li, F. Automatic semantic annotation for image retrieval based on multiple kernel learning. In *Proceedings of the International Conference on Logistics, Engineering, Management and Computer Science* (2014), pp. 649–653.

[32] Isabelle, J. Semantic, automatic image annotation based on multi-layered active contours and decision trees. *International Journal of Advanced Computer Science and Applications 4* (2013), 201–208.

[33] Joachims, T. Text categorization with support vector machines: Learning with many relevant features. In *Machine Learning: ECML-98: 10th European Conference on Machine Learning* (2005).

[34] Kitchenham, B. Procedures for performing systematic reviews. *Keele University 33* (2004), 1–26.

[35] Kitchenham, B. Guidelines for performing systematic literature reviews in software engineering. Technical Report Keele University and Durham University Joint Report, 2007.

[36] Kurdi, G. Toward an electronic resource for systematic reviews in computer science, 2022. Researchgate.net.

[37] Körner, D. . Automated semantic annotation of historical catalogues. Master thesis, 2020.

[38] Le, H., Nguyen, M., and Yan, W. Machine learning with synthetic data - a new way to learn and classify the pictorial augmented reality markers in real-time. In *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)* (2020), pp. 1–6.

[39] Li, R., and Li, S. Multimedia image data analysis based on knn algorithm. In *Computational Intelligence and Neuroscience* (2022), p. 7963603.

[40] Liao, X., and Zhao, Z. Unsupervised approaches for textual semantic annotation, a survey. *ACM Computing Surveys 52* (2019), 1–45.

[41] Lin, S., Chung, C., Hu, W., Hung, C., Chen, S., and Lin, T. Automated knowledge discovery and semantic annotation for network and web services. *International Journal of Distributed Sensor Networks 12* (2016), 1550147716657925.

[42] Lin, S., Li, J., and Yu, C. Dynamic data driven-based automatic clustering and semantic annotation for internet of things sensor data. *Sensors and Materials 31* (2019), 1789–1801.

[43] Liu, F., Cui, J., Janssens, D., Wets, G., and Cools, M. Semantic annotation of mobile phone data using machine learning algorithms. *Smartphones from an Applied Research Perspective* (2017).

[44] Liu, F., Li, P., and Deng, D. Device-oriented automatic semantic annotation in iot. *Journal of Sensors 2017* (2017), 1–14.

[45] Liu, Z., Bao, J., and Ding, F. An improved k-means clustering algorithm based on semantic model. In *International Conference on Information Technology and Electrical Engineering* (2018), pp. 1–5.

[46] Mahdavi-Hezavehi, D., Galster, M., and Avgeriou, P. Variability in quality attributes of service-based software systems: A systematic literature review. *Information and Software Technology 55* (2013), 320–343.

[47] Makris, C., and Simos, M. Otnel: A distributed online deep learning semantic annotation methodology. *Big Data and Cognitive Computing 4* (2020), 31.

[48] Marbrouk, C., and Konaté, K. An approach to extracting distributed data from the integrated environment of web technologies based on set theory. *International Journal of Computer Science and Information Technology 11* (2019), 29–44.

[49] Mesbah, S., Fragkeskos, K., Lofi, C., Bozzon, A., and Houben, G. Semantic annotation of data processing pipelines in scientific publications. In *The Semantic Web: 14th International Conference, ESWC 2017* (2017), pp. 321–336.

[50] Miranda, P., Isaias, P., and Costa, C. E-learning and web generations: Towards web 3.0 and e-learning 3.0. In *International Proceedings of Economics Development and Research, IPEDR* (2014), pp. 92–103.

[51] Mwiti, D. 10 real-life applications of reinforcement learning, 2023. https://neptune.ai/blog/reinforcement-learning-applications.

[52] Patra, A., and Singh, D. A survery report on text classification with different term weighing methods and comparison between classification algorithms. *International Journal of Computer Applications 75* (2013), 14–18.

[53] Pech, F., Martinez, A., Estrada, H., and Hernandez, Y. Semantic annotation of unstructured documents using concepts similarity. *Scientific Programming 2017* (2017), 1–10.

[54] Raj, R. Supervise, unsupervised, and semi-supervised learning with real-life use case, 2020. www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning.

[55] Rinaldi, F. Semi-automated semantic annotation of the biomedical literature. In *ISWC (Posters & Demos)* (2014), pp. 473–476.

[56] Salleh, N., Mendes, E., and Grundy, J. Empirical studies of pair programming for CS/SE

[85] Zakharova, O. Main aspects of big data semantic annotaion. *Problems in Programming 4* (2020), 022–033.

[86] Zhang, J. Vision to keywords: Automatic image annotation by filling the semantic gap. Doctoral dissertation, 2019.

[87] Zhang, J., Wen, X., Cho, A., and Whang, M. An empathy evaluation system using spectogram image features of audio. *Sensors 21* (2021), 7111.

[88] Zhang, M., Han, L., Yuan, L., and Chen, N. Ontology-based automatic semantic annotation method for iot data resources. In *2020 International Conferences on Internt of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)* (2020), pp. 661–667.

[89] Zhang, P., Du, J., Fan, D., and Zhou, Y. Automatic image semantic annotation based on the tourism domain ontological knowledge base. In *Communications in Computer and Information Science* (2015), pp. 61–69.