



The Evaluation of Ontologies for Quality, Suitability for Reuse, and the Significant Role of Social Factors

Marzieh Talebpour^(✉), Martin Sykora, and Tom Jackson

School of Business and Economics, Loughborough University, Loughborough, UK
{m.talebpour, m.d.Sykora, t.w.Jackson}@lboro.ac.uk

Abstract. A frequent challenge faced by ontologists and knowledge engineers is the choice of the correct or most appropriate ontology for reuse. Despite the importance of ontology evaluation and selection and the widespread research on these topics, there are still many unanswered questions and challenges. Most of the evaluation metrics and frameworks in the literature are mainly based on a limited set of internal characteristics of ontologies, e.g., their content and structure, which ignore how the community uses and evaluates them. This paper used a survey questionnaire to investigate the notion of quality and reusability in ontology engineering, and to explore and identify the set of metrics that can affect the process of ontology evaluation and selection for reuse. Responses from 157 ontologists and knowledge engineers were collected, and their analysis suggests that the process of ontology evaluation and selection for reuse, not only depends on different internal characteristics of ontologies, but that it also depends on different metadata, and social and community related metrics. Findings of this research can contribute to facilitating and improving the process of selecting an ontology for reuse.

Keywords: Quality metrics · Ontology · Evaluation · Ontology selection · Ontology reuse

1 Introduction

Recent uptake in Semantic Web technology applications has urged researchers and ontology engineers to develop ontologies in different domains. Increase in the number of ontologies and the cost of developing them has urged researchers in this field to consider ontology reuse [1]. Ontology reuse can be defined as the process of using the available ontological knowledge as input to develop new ontologies. Building an ontology by reusing the available ones will not only facilitate the development process but will also make the outcome ontology reusable. Ontology reuse consists of different steps namely searching for adequate ontologies, evaluating the quality and fitness of those ontologies for the reuse purpose, selecting an ontology and integrating it in the project [2].

Regardless of all the advantages of reusing ontologies and the availability of different ontologies, ontology reuse has always been a challenging task. Guidelines for building ontologies are usually blamed for lack of reuse strategies and some argue that they are

not explicitly concerned with ontology reuse. Others consider the first steps of ontology reuse, that is the identification and evaluation of the knowledge sources which can be useful for an application domain, as the hardest step in the process of ontology reuse. Ontologists and knowledge engineers not only have to find the most appropriate ontologies for their search query but should also be able to evaluate those ontologies according to different implicit or explicit criteria. The lack of appropriate supportive tools and automatic measurement techniques for evaluating and assessing ontology features has been considered as a barrier for ontology reuse [3].

Ontology evaluation is at the heart of ontology selection and has received a considerable amount of attention in the literature. The term evaluation refers to the process of judging different technical aspects of an ontology namely its definitions, documentation and software environment [4]. Evaluation has also been described as the process of measuring the suitability and the quality of an ontology for a specific goal or in a specific application [3]. This definition refers to the approaches that aim to identify an ontology, an ontology module or a set of ontologies that satisfy a particular set of selection requirements [5].

This paper is an extended version of [6], and aims to determine some of the metrics that can be used to evaluate the suitability of an ontology for reuse. The fundamental research question of this study was whether or not social and community related metrics can be used in the evaluation process. Another question was how important those metrics were, compared to some of the well-known ontological metrics such as content and structure. Qualitative and quantitative research designs were adopted to provide a deeper understanding of how ontologists and knowledge engineers evaluate and select ontologies. This study offers some valuable insights into ontology quality, what it depends on and how it can be measured.

2 Background

Evaluation is one of the most popular and also defined terms in the field of ontology engineering. It is used to refer to several different activities including detecting faults in an ontology, assessing an ontology's quality, and measuring its fitness for a specific purpose. There are many different ways of defining ontology evaluation; one of the most popular and also the earliest definitions for ontology evaluation was provided by Gómez-Pérez where the term evaluation was used to refer to the technical judgment of an ontology considering its different aspects, namely its definitions, documentation, and software environment [4]. According to this definition, evaluation encompasses validation and verification; ontology validation is mainly concerned with the correctness of an ontology whereas ontology verification is more about determining how well an ontology corresponds to what it should represent [7]. In other words, ontology validation focuses on building the correct ontology whereas ontology verification is about building an ontology correctly [8].

Ontology evaluation has also been widely defined as the process of determining the adequacy and quality of an ontology for being used for a specific goal and in a specific context [3]. This definition is used to link the process of ontology evaluation to ontology selection. Ontology selection aims to identify an ontology, an ontology module or a set

of ontologies that satisfy a particular set of criteria or selection requirements [5]. Some consider ontology evaluation as the core to ontology selection and argue that ontology evaluation is influenced by different components of the selection process, e.g., selection criteria, type of output, and the libraries that the selection is based on [5]. Ontology assessment is also used to refer to this particular definition of ontology evaluation and is commonly defined as the activity of checking and judging an ontology against different user requirements such as usability and usefulness [9]. Unlike the first definition of the ontology evaluation, in which the developer team is responsible for validating and verifying an ontology, ontology assessment and evaluation for selection is done by the end users [10].

Ontology evaluation can also refer to a function or an activity that aims to map an ontology or a component of an ontology to a score or a number, e.g., in the range of 0 to 1 [11]. The main aim of these types of processes is to measure and assess the quality of an ontology with regards to a set of predefined metrics and requirements [12]. This definition is somehow similar to what [9] defines as ontology quality assurance, which refers to the activity of examining every process carried out and every product built during the ontology development process and making sure that the level of their quality is satisfactory. Moreover and as it is seen in the literature, the expressions “Ontology Evaluation” and “Ontology Ranking” are sometimes used interchangeably. While they both tend to refer to a set of similar criteria, for us, ontology ranking is the process of sorting ontologies in descending order and according to the scores that are assigned to them in the evaluation process.

Ontology evaluation is important in the ontology development process, whether it is built from scratch, automatically or by reusing other ontologies [13]. While building an ontology from scratch, developers need to evaluate the outcome ontology, to measure its quality [14], to check if it meets their application requirements [13] and also to identify the potential refinement steps [15]. Evaluation is also helpful in checking the homogeneity and consistency of an ontology when it is automatically populated from different resources [13, 16]. Building an ontology from scratch is very costly and time-consuming [17, 18]; therefore, ontologists are urged to consider reusing existing ontologies before building a new one [19]. Ontology evaluation is and has always been a major concept when it comes to ontology reuse [20]. Some argue that ontology evaluation is one of the main issues that should be addressed if ontologies are to become widely adopted and reused by the community [15, 18, 20, 21].

Moreover, the number of ontologies on the web has been increasing rapidly [13], and users usually face multiple ontologies when they need to choose or use one in their everyday activities [12, 15, 22]. Before using an ontology in an application or selecting it for reuse, ontologists have to assess its quality and correctness and also compare it to the other available ones in the domain. This is when ontology evaluation comes into the picture; ontology evaluation is believed to be the core to the ontology selection process [5] and is used to select the best or the most appropriate ontology among many other candidates in a domain [15]. Evaluating an ontology is considered as a complicated process [12, 23]; it is believed that failure to evaluate ontologies or to choose the right ontology can lead to using the ontologies that are not right or have a lower quality [12].

Being one of the most popular and also important parts of the ontology engineering domain, ontology evaluation has long been at the centre of research attention in this field. Since 1995 to date, there has been a variety of research on different aspects of ontology evaluation including methodologies, tools, frameworks, methods, metrics, and measures [4]. However, much uncertainty and also disagreement still exists about the best way to evaluate an ontology generally or for a specific tool or application. As it is seen in the literature, there are many different ways of evaluating ontologies and also many ways of classifying those evaluation methods, algorithms and approaches. Some of the most popular ontology evaluation approaches are reviewed in the following part of this section. Ontology evaluation approaches can broadly be classified as follow:

User-Based Evaluation. Ontologists and knowledge experts can assess the quality of ontologies [8] in two different ways: one is the criteria-based evaluation approach in which the suitability of an ontology for a particular task or requirement is evaluated by being compared against a set of pre-defined criteria [18]. Peer review based evaluation, as the other type of user-based evaluation approach, allows ontologists and knowledge experts to link subjective information to ontologies by providing metadata and extra qualitative information about different aspects of them [24]. Despite their popularity, user-based ontology evaluation approaches are blamed for being solely based on different characteristics of ontologies and for ignoring the functionality of an ontology in an application [12].

Golden Standard. This approach refers to the type of evaluation that is performed by comparing an ontology to another ontology, also known as a “gold standard” ontology, and aims to find different types of similarities such as lexical as conceptual between them. This approach was first proposed by [25] and was then used in many other researches namely [11], where a fully automated evaluation approach was proposed by introducing a similarity measure called OntoRand index and comparing ontologies to a gold standard one using that measure. This kind of evaluation is typically applied to the ontologies that are generated semi-automatically and to measure the effectiveness of the ontology generation process [22]. A major problem with this approach is that comparing ontologies is not easy [5].

Data or Corpus Driven Evaluation. This approach is similar to the “gold standard” approach, but instead of comparing an ontology to another one, it compares it to a source of data or a collection of documents [15]. One of the most popular architectures for this type of evaluation is proposed by [19]; it is based on three main steps namely extracting keywords from a corpus, applying some query expansion algorithms on the ontology concept, and finally mapping the terms identified in the corpus to the concepts in an ontology. They will then analyse how well the ontology is covering the source of data [19].

Task-Based Evaluation. Also known as application-based [26] or black box evaluation [21]; this approach aims to evaluate an ontology’s performance in the context of an application [19]. One of the main assumptions of this approach is that there is a direct link between the quality of an ontology and how well it serves its purpose as a part of a larger application [27]. The challenges of performing this type of evaluation includes

the difficulty of assessing the quality of the performed task as well as making sure that the experimental environment is clean, and that the ontology is the only factor that is influencing the performance of the application [5].

Rule-Based (Logical). This type of evaluation is proposed by [16] and aims to validate ontologies and detect conflicts in them by using different rules that are either a part of the ontology development language or are identified by users. Rule-based evaluation is more relevant when evaluation aims to detect faults and inconsistencies in an ontology, rather than when the quality assessment or ontology selection is concerned.

Other Approaches. Besides the above-mentioned categories, that are very popular in the literature, there are some other ways of classifying ontology evaluation approaches. For example, ontology evaluation approaches can be classified into glass-box or black-box. Glass-box approaches tend to evaluate the internal content and structure of ontologies [20] and are blamed for not predicting how ontology might perform in an application. In contrast, black-box approaches do not explicitly use knowledge of the internal structure of ontologies and focus on the quality of an ontology performance and results [20]. Ontologies can also be evaluated as a whole or according to their different layers, e.g. data level, taxonomy level, and application level [15]. [17] has divided the concept of ontology quality into two broad types: “Total Quality” and “Partial Quality”. Some argue that evaluating an ontology as a whole, especially automatically, is not possible or practical, especially considering the complex structure of ontologies [15].

From all the approaches mentioned above, much of the research in the ontology evaluation domain has concentrated on criteria-based approaches, and many have tried to identify and introduce a set of metrics that can be used for ontology evaluation. A more detailed account of criteria-based ontology evaluation is given in the next section.

3 Criteria-Based Evaluation

Criteria-based evaluation, also known as metric-based, multiple-criteria [15] or feature-based [16], is one of the most popular evaluation approaches in the literature. This type of evaluation is mostly based on identifying and selecting multiple attributes or features of ontologies and then evaluating them for ranking and selection purposes [15]. The outcome of this approach is usually an overall or an aggregated score that is computed by adding the scores that are assigned to each criterion [28]. Despite the wide use and popularity of criteria-based evaluation, identifying the right set of metrics for ontology evaluation and measuring them is still a challenge.

Criteria based approaches are different from each other in a number of respects. First, the type of the metrics they use to assess ontologies can be different. Some approaches are based on qualitative metrics and tend to rely on expert users’ judgement and ratings about an ontology or a module in an ontology [29]. Qualitative approaches can also be used to evaluate an ontology based on the principles that are/were used in its construction [19]. Other are based on different quantitative criteria about different aspects of ontologies such as its structure and content. These approaches, that are also known as formal rational

approaches, are usually concerned with technical and economical aspects of ontologies and use different goal-based strategies [18].

They can also be based on assessing internal and/or external attributes of ontologies. Internal attributes are concerned with the ontology itself and its internal organization whereas external measures mostly focus on how ontologies are taken-up or used within the user communities [30]. [31], for example, has followed software engineering measurement traditions and has proposed a method that aims to identify what they call key internal attributes of ontologies including consistency, richness and clarity. They have also mentioned maintainability and application performance as example for external quality attributes of ontologies [31].

Moreover, metrics used in the criteria-based evaluation can either be query dependent or query independent. Coverage, for example, aims to measure how well a candidate ontology match or cover a set of query term(s) and selection requirements [32, 33] and therefore, it depends on users' queries. Popularity, in contrast, is measured by checking the presence of an ontology in different well-known repositories as well as looking into the number of visits or page views of an ontology in ontology repositories in a recent specific period [28]; hence, it does not depend on the selection requirements.

For the purpose of this paper and according to the previous study conducted by [34], ontology evaluation quality criteria are broadly classified into three main sub groups including (1) Internal metrics that are based on different internal characteristic of ontologies such as their content and structure, (2) Metadata that are used to describe ontologies and to help in the selection process, and (3) Social metrics that focus on how ontologies are used by communities. The rest of this section moves on to explain different quality metrics for ontology evaluation in more details.

3.1 Internal Metrics

Internal aspects of ontologies have always been used as a mean of their evaluation. Different internal quality criteria such as clarity, correctness, consistency, and completeness have been used in the literature to measure how clear ontology definitions are, how different entities in an ontology represent the real world, how consistent an ontology is, and how complete an ontology is [12]. Coverage is yet another significant content related metric; the term coverage is mostly used in the literature to measure how well a candidate ontology match or cover the query term(s) and selection requirements [32]. Structure or graph structure [20] is the other important internal aspect of an ontology that can be used to measure how detailed the knowledge structure of an ontology is [35] and also to evaluate its richness of knowledge [5] density [22], depth and breadth [35].

3.2 Metadata

Besides the internal aspects of ontologies, some of the frameworks and tools have suggested evaluating ontologies using different types of metadata. Metadata or "data about data" is widely used on the web for different reasons namely to help in the process of resource discovery [36]. [37] believes that the primary connection between different elements of an ontology is in the mind of the people who interpret it; so, tagging an ontology with more data will help in making those mental connections explicit. Ontologies can

be tagged and described according to their different characteristics, namely their type and version. The language that different ontologies are built and implemented with can also be used as a metric to evaluate, filter and categorize them [38].

There are different examples of using metadata in the literature to help with the process of evaluating, finding and reusing ontologies. Swoogle [39] was one of the very first selection systems in ontology engineering field to introduce the concept of metadata to this domain. There is a metadata generator component in this system that is responsible for creating and storing three different types of metadata about each discovered ontology including basic, relation, and analytical metadata [39]. [24] has also proposed two sets of metadata that can be used to evaluate ontologies: source metadata and third-party metadata.

Moreover, metadata is created and used to help interoperability between different applications and ontologies. Ontology Metadata Vocabulary (OMV) was proposed by [40] and is one of the most popular sets of metadata for ontologies. OMV is not directly concerned with ontology evaluation or ranking and its main aim is to facilitate ontology reuse. [41] have proposed a guideline for minimum information for the reporting of an ontology (MIRO) to help ontologists and knowledge engineers in the process of reporting ontology description and providing documentation. It is believed that MIRO can improve the quality and consistency of ontology descriptions and documentation.

3.3 Community Aspects of Ontologies

Besides how ontologies are built and what they are covering or even not covering, some believe that how they are used by different communities can be considered as a feature in their evaluation and selection. [8] define user-based ontology evaluation as the process of evaluating an ontology through users' experiences and by capturing different subjective information about ontologies. According to a study that was conducted by [42], relying on the experiences of other users for evaluating ontologies will lessen the efforts needed to assess an ontology and reduce the problems that users face while selecting an ontology. [23] have also highlighted the importance of relying on the wisdom of the crowd in ontology evaluation and believe that improving the overall quality of ontological content on the web is a shared responsibility within a community.

As it is seen in the literature, social or community features of ontologies have not been the main focus of the evaluation frameworks until recently. However, some of the very well-known frameworks for ontology evaluation consider social quality as one of the metrics, among others, that can be used in the evaluation process. [31], for example, applied a deductive method to identify a set of general, domain-independent and application-independent quality metrics for ontology evaluation. This approach proposed different social quality metrics namely authority and history to measure the role of community in ontology quality.

Another example of social based quality application was proposed by [43], in which the notion of the open rating system and democratic ranking were applied to ontology evaluation. According to this approach, users of this system can not only review the ontology, but they can also review the reviews provided by other users about an ontology. A similar approach was proposed by [42] where users' ratings are used to determine what they call user-perceived quality of ontologies.

[34] also attempted to investigate and explore how community and social aspects of ontologies can affect their quality. According to their findings, knowledge engineers consider different social aspects of ontologies when evaluating them. Those aspects include: (1) build related information, for example, who has built the ontology, why the ontology was built, do they know the developer team, (2) regularity of update and maintenance, and (3) responsiveness of the ontology developer and maintenance team and their flexibility and willingness toward making changes.

Overall, the above-mentioned studies highlight the importance of the criteria-based approaches in ontology evaluation. They also outline the most important or used quality metrics in the literature. The next sections discuss the methodology used to collect data and the findings of this research.

4 Methodology

From all the groups of quality related metrics mentioned in the previous section, the focus of this research is on different metadata and social characteristics of ontologies that can be used in the evaluation process. This study was built upon the findings of the previous interview study conducted by [34] and aims to clarify and confirm the metrics identified in that study. To do that a survey questionnaire was designed based on a mixed research strategy combining qualitative and quantitative questions.

The survey was sent to a broad community of ontologists and knowledge engineers in different domains. Different sampling strategies namely purposive sampling [44] were used in order to find the ontologists and knowledge engineers that were involved in the process of ontology development and reuse. The survey was also forwarded to different active mailing lists in the field of ontology engineering. The lists used are as follows:

- The UK Ontology Network
- GO-Discuss
- DBpedia-discussion
- The Protégé User
- FGED-discuss
- Linked Data for Language Technology Community Group
- Best Practices for Multilingual Linked Open Data Community Group
- Ontology-Lexica Community Group
- Linking Open Data project
- Ontology Lookup Service announce
- Technical discussion of the OWL Working Group
- This is the mailing list for the Semantic Web Health Care and Life Sciences Community Group

There was a total number of 31 questions broadly divided into four different sections. Each section consisted of different number of questions and aimed to explore and discover the opinion of ontologists and knowledge engineers regarding (1) the process of ontology development, (2) ontology reuse, (3) ontology evaluation and the quality metrics used in that process, and (4) the role of community in ontology development, evaluation and reuse. Different types of questions were used in the survey namely close-ended

questions, Likert scale questions, open-ended questions, and multiple-choice questions. Screening questions were also used throughout the survey to make sure that respondents are presented with the set of questions that is relevant to their previous experiences.

The most important part of the survey aimed to explore the process of ontology evaluation and the set of criteria that can be used in this process. Respondents were first asked about the approaches and metrics they tend to consider while evaluating ontologies. They were then presented with four different sets of quality metrics including (1) internal, (2) metadata, (3) community and (4) popularity related criteria and were asked how important they thought those metrics were, by offering a 5-point Likert scale, ranging from “Not important” to “Very important”. The criteria presented and assessed in this part of the survey were collected both from the literature and the previous phase of the data collection, that was an interview study with 15 ontologists and knowledge engineers in different domains [34].

5 Findings

As was mentioned in the previous sections, this research aimed to introduce different metrics that could be potentially used for ontology evaluation. Prior studies have identified many different quality metrics, mostly based on ontological and internal aspects of ontologies. This study was designed to determine the importance of those metrics and also to explore how communities can help in the selection process. The findings of this study are discussed in the following sections.

5.1 Demographics of Respondents

The aim of this section is to provide information on the profile of respondents to the survey. This study managed to access ontologists and knowledge engineers with many years of experience in building and reusing ontologies in different domains. Around 80% of the participants in the survey were actively involved in the ontology development process and all of them would consider reusing existing ontologies before building a new one. The 157 respondents of this study are categorized by the following demographics, all declared by responders:

Job Title. After conducting frequency analysis on the job titles provided by respondents, 78 unique job titles were identified, many of which were somehow related to different roles and positions in academia such as researcher, professor, lecturer, etc.

Type of Organization. According to the frequency analysis conducted on the organization types, 68.8% (108) of the respondents of the survey were working in academia. The other 31.2% of the respondents were working in other types of organizations including different companies and industries.

Years of Experience. Interestingly, most of the survey respondents were experts in their domain and only around 10% of them had less than two years of experience. Around 46% (73) of the respondents had more than ten years of experience. The second largest group of the respondents were the ontologists with five to ten years of experience (26.8%).

Main Domains They Had Built or Reused Ontologies In. Survey respondents had worked/were working in many different domains such as biomedical, industry, business, etc. Most of participants had mentioned more than one domain, some of which were not related to each other.

5.2 Evaluation Metrics According to Qualitative Data

Before presenting participants with four sets of quality metrics that can be used for ontology evaluation and asking them to rate those metrics, they were asked an open-ended question about how they evaluate the quality of an ontology before selecting it for reuse. This question aimed to provide further insight and to gather respondents' opinions on different evaluation metrics and approaches. The responses to this question were coded according to different categories of quality metrics namely (1) internal, (2) metadata, (3) community and popularity related metrics.

According to the analysis, quality metrics thought to be the most important were content and coverage (mentioned 51 times) and documentation (mentioned 41 times). The fact that an ontology has been reused previously and the popularity of the ontology on the web, or among community was the other frequently mentioned metric by the respondents (38 times). Community related metrics such as reviews about the quality of an ontology, existence, activeness and responsiveness of the developer team, and the reputation of the developer team or organisation responsible for ontology were also mentioned by many of the respondents (25 times).

The findings of the qualitative question in the survey confirmed the findings of the quantitative part and the interview study previously conducted by [34]. It should be noted that two of the metrics mentioned by the responders namely "fit" and "format" were not presented as a Likert item in the quantitative part of the survey. Format was only mentioned two times but how relevant an ontology is to an application requirement was mentioned 37 times. The reason fit was not used as a Likert item is that it cannot be used as a criterion to judge the quality of an ontology. However, it is a significant factor in the selection process.

One of the emerging themes in the analysis was "following or being a part of a standard". Interestingly, 19 respondents had mentioned following or complying with different design guidelines and principles or being a part of a standard like W3C, and OBO Foundry as a criterion in the evaluation process. Some had also mentioned that while evaluating an ontology, they check if it is built by using a method like NEON. A similar question was proposed as one of the Likert items and respondents were asked to rate how important "The use of a method/methodology (e.g. NEON, METHONTOL-OGY, or any other standard and development practice)" is when evaluating an ontology. Surprisingly, it was ranked 30th (out of 31) with a mean of 2.80 and a median of 3.

5.3 Importance of Quality Metrics

Table 1 shows the descriptive statistics of all 31 quality metrics, sorted by standard deviation. The metrics are ranked from 1 to 31, with 1 being the most important and 31 being the least important metric considered when evaluating the quality of an ontology

for reuse. Mean and median are used to show the center and midpoint of the data respectively. Standard deviation is used to express the level of agreement on the importance of each metric in the ontology evaluation process; the lower value of standard deviation represents the higher level of agreement among the survey respondents on a rating.

As it is seen in Table 1, ontology content including its classes, properties, relationships, individuals and axioms is the first metric ontologists and knowledge engineers tend to look at when evaluating the quality of an ontology for reuse. Other internal aspects of ontologies like their structure (class hierarchy or taxonomy), scope (domain coverage), syntactic correctness, and consistency (e.g. naming and spelling consistency all over the ontology) are also among the top ten quality metrics used for ontology evaluation.

Documentation is the second most important quality metric used in the evaluation process. Survey respondents have also given a very high rate, five and eight respectively, to other metadata related metrics such as accessibility and availability of metadata and provenance information about an ontology. In contrast to these metrics, other criteria in the metadata group like availability of funds for ontology update and maintenance, use of a method/methodology and ontology language are among the bottom ten least important metrics.

Community related metrics have some very interesting ratings. The results show ontologists and knowledge engineers would like to know about the purpose that an ontology is used/has been used for (e.g. annotation, sharing data, etc.) while evaluating and before selecting it for reuse. They have also rated “Availability of wikis, forums, mailing lists and support team for the ontology” as one of the very important quality metrics for ontology evaluation. Having an active, responsive developer community and knowing and trusting the ontology developers are among the other top-ranked community related aspects of ontologies that can be used for their evaluation.

Survey responders were also presented with a set of popularity related metrics. According to Table 1, the popularity of an ontology in the community and among colleagues has the highest median and mean compared to the other metrics that can be used for evaluating the popularity of an ontology. Respondents also tended to consider the reputation of the ontology developer team and/or institute in the domain while evaluating an ontology for reuse. Other popularity related metrics such as the popularity of the ontology in social media (e.g. in GitHub, Twitter, or LinkedIn), the popularity of the ontology on the web (number of times it has been viewed in different websites/applications across the web), and the reviews of the ontology (e.g. ratings), were among the metrics with the least mean and median.

Survey responders were also presented with a set of popularity related metrics. According to Table 1, the popularity of an ontology in the community and among colleagues has the highest median and mean compared to the other metrics that can be used for evaluating the popularity of an ontology. Respondents also tended to consider the reputation of the ontology developer team and/or institute in the domain while evaluating an ontology for reuse. Other popularity related metrics such as the popularity of the ontology in social media (e.g. in GitHub, Twitter, or LinkedIn), the popularity of the ontology on the web (number of times it has been viewed in different websites/applications across the web), and the reviews of the ontology (e.g. ratings), were among the metrics with the least mean and median.

Table 1. Descriptive statistics of all the quality metrics in the survey (extracted from [6]).

Rank	Metric	SD	Median	Mean
1	The Content (classes, properties, relationships, individuals, axioms)	0.57	5	4.59
2	The availability of documentation (both internal, e.g. adding comments and external)	0.79	5	4.38
3	The Structure (Class hierarchy or taxonomy)	0.82	4	4.29
4	The Scope (domain coverage)	0.84	5	4.42
5	The ontology is online, accessible, and open to reuse (e.g. License type)	0.85	5	4.52
6	The Syntactic Correctness	0.92	4	4.15
7	The Consistency (e.g. Naming and spelling consistency all over the ontology)	1.00	4	4.03
8	Availability of metadata and provenance information about the ontology	1.01	4	3.92
9	Availability of wikis, forums, mailing lists and support team for the ontology	1.03	4	3.45
10	Having information about the purpose that ontology is used/has been used for (e.g. annotation, sharing data, etc.)	1.03	4	3.77
11	The Semantic Richness and Correctness (e.g. level of details)	1.06	4	3.92
12	Having an active responsive (developer) community	1.09	4	3.62
13	Having information about the other individuals or organizations who are using/have used the ontology	1.1	3	3.12
14	Having information about the other projects that the ontology is used/has been used in	1.1	3	3.34
15	Knowing and trusting the ontology developers	1.11	4	3.42
16	Knowing and trusting the organization or institute that is responsible for ontology development	1.11	3	3.38
17	The reputation of the ontology developer team, and/or institute in the domain	1.12	3	3.31
18	The number of times the ontology has been reused or cited (e.g. owl:imports, rdfs:seeAlso, daml:sameClassAs)	1.13	3	3.40
19	The flexibility of the Ontology (being easy to change) and the ontology developer team	1.14	4	3.41
20	The frequency of updates, maintenance, and submissions to the ontology	1.16	3	3.22
21	The popularity of the ontology in social media (e.g. in GitHub, Twitter, or LinkedIn)	1.16	2	2.28

(continued)

Table 1. (continued)

Rank	Metric	SD	Median	Mean
22	The popularity of the ontology in the community and among colleagues	1.17	4	3.51
23	The number of updates, maintenance, and submissions to the ontology	1.19	3	3.13
24	Availability of published(scientific) work about the ontology	1.19	4	3.56
25	The size of the ontology	1.19	3	3.02
26	The number of times the ontology has been reused or cited (e.g. owl:imports, rdfs:seeAlso, daml:sameClassAs)	1.19	3	3.08
27	The availability of funds for ontology update and maintenance	1.23	3	2.77
28	The popularity of the ontology on the web (number of times it has been viewed in different websites/applications across the web)	1.24	3	3.05
29	The reviews of the ontology (e.g. ratings)	1.25	3	3.03
30	The use of a method/methodology (e.g. NEON, METHONTOLOGY, or any other standard and development practice)	1.26	3	2.80
31	The Language that ontology is built in (e.g. OWL)	1.30	4	3.70

6 Discussion

Finding a set of metrics that can be used for ontology evaluation and selection for reuse has always been a critical research topic in the field of ontology engineering. As mentioned in the introduction and background sections, many different ontology evaluation approaches and metrics for quality assessment have been proposed in the literature. However, these studies suffer from some limitations; for example, they have not dealt with ranking and the importance of the quality metrics, especially the community related ones. Therefore, the focus of this research was on constructing a criteria-based evaluation approach and determining a set of metrics that ontologists and knowledge engineers tend to look at before selecting an ontology for reuse. This study also set out with the aim of assessing the importance of the quality metrics identified in the literature and in a previous phase of this research [34].

Previous studies have mostly been concerned with identification and application of a new set of quality metrics [38]. However, the key aim of this study was not only to identify the quality metrics used in the process of evaluating ontologies but also to find how important each of the quality metrics are. The results of this survey study indicate that the internal characteristics of ontologies are the first to assess before selecting them for reuse. However, some other aspects of ontologies such as availability of documentation, availability and accessibility of an ontology (e.g. license type), availability of metadata and provenance information, and also having information about the purpose that ontology

is used/has been used for previously (e.g. annotation, sharing data, etc.) are as important as the quality of the internal components of ontologies.

Popularity, as one of the most defined and used term in the literature, refers to the role of community in the quality assessment process. As a part of this study, respondents were asked to rate the importance of six different popularity related metrics, four of which were previously mentioned in the literature. The results suggest that ontologists and knowledge engineers tend to care more about the popularity metrics, as identified by [34, 45], such as popularity of an ontology in the community and among colleagues (ranked 14 out of 31, when sorted by median) and the reputation of the ontology developer team, and/or institute in the domain (ranked 21 out of 31, when sorted by median) than the popularity related metrics that have been widely used in the literature and by selection systems. Metrics used in the literature include the number of times an ontology has been reused or cited [46, 47], the popularity of an ontology on the web [28, 31], the reviews of an ontology [42] and the popularity of an ontology on social media [48]; while having a lower median and mean, some of these metrics were ranked higher when the quality metrics were sorted by standard deviation. Standard Deviation shows a higher level of agreement among the survey respondents about the lower rank of those metrics.

7 Conclusion

This paper set out to explore and clarify the notions of quality and reuse in the field of ontology engineering and to identify the set of metrics that ontologists and knowledge engineers tend to consider when assessing the suitability of an ontology for reuse. It also investigated the potential role of community and social interactions in the process of ontology evaluation and selection for reuse.

The results of this study suggest that the process of ontology evaluation and selection for reuse does not only depend on different internal characteristics of ontologies, such as their content and structure, but it also depends on many other metadata and community related metrics. Moreover, the results of this study indicate that ontologists and knowledge engineers find some of the metrics identified in this research more important and useful, compared to the ones proposed by the previous studies. The proposed ranking based on the metrics identified in this research were also found helpful and useful in the ontology evaluation and selection process.

Overall, the results suggest that the metadata and social related metrics should be used by different selection systems in this field in order to facilitate and improve the process of evaluating and selecting ontologies for reuse and also, to provide a more comprehensive and accurate recommendation for reuse. Moreover, definition of some of the quality metrics used in the literature, e.g., popularity and how they are currently measured, may benefit from updating.

References

1. Bontas, E.P., Mochol, M., Tolksdorf, R.: Case studies on ontology reuse. In: Proceedings of the IKNOW05 International Conference on Knowledge Management, vol. 74, p. 345, June 2005

2. d'Aquin, M., et al.: What can be done with the semantic web? An overview of watson-based applications. In: CEUR Workshop Proceedings, vol. 426 (2008)
3. Fernández, M., Cantador, I., Castells, P.: CORE: a tool for collaborative ontology reuse and evaluation. In: CEUR Workshop Proceedings, vol. 179 (2006)
4. Gómez-Pérez, A.: Some ideas and examples to evaluate ontologies. In: 11th Conference on Artificial Intelligence for Applications Proceedings, pp. 299–305. IEEE, February 1995
5. Sabou, M., et al.: Ontology selection: ontology evaluation on the real semantic web. In: Vrandečić, D., et al. (ed.) 4th International EON Workshop, Evaluation of Ontologies for the Web, EON 2006, (CEUR Workshop Proceedings), Edinburgh, Scotland, UK (2006)
6. Talebpour, M., Sykora, M.D., Jackson, T.: Ontology selection for reuse: will it ever get easier?. In: 10th International Conference on Knowledge Engineering and Ontology Development, 18–20 September, Seville, Spain (2018)
7. Gómez-Pérez, A., Facultad, D.I.: Evaluation of taxonomic knowledge in ontologies and knowledge bases (1999). <http://sern.ucalgary.ca/KSI/KAW/KAW99>
8. Hloman, H., Stacey, D.: Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: a survey. *Semant. Web J.* **1**, 1–11 (2014). <http://www.semantic-web-journal.net/system/files/swj657.pdf>
9. Suárez-Figueroa, M.C., Gómez-Pérez, A.: First attempt towards a standard glossary of ontology engineering terminology. In: 8th International Conference on Terminology and Knowledge Engineering, TKE 2008 (2008). <http://www.neon-project.org/>. Accessed 17 Feb 2019
10. Gómez-Pérez, A.: From knowledge based systems to knowledge sharing technology: evaluation and assessment, differences, pp. 1–15 (1994). <http://oa.upm.es/6498/>
11. Brank, J., Mladenic, D., Grobelnik, M.: Gold standard based ontology evaluation using instance assignment. In: Workshop on Evaluation of Ontologies for the Web, EON, May 2006
12. Yu, J., Thom, J.A., Tam, A.: Requirements-oriented methodology for evaluating ontologies. *Inf. Syst.* **34**(8), 766–791 (2009)
13. Tartir, S., Arpinar, I.B., Sheth, A.P.: Ontological evaluation and validation. In: Poli, R., Healy, M., Kameas, A. (eds.) *Theory and Applications of Ontology: Computer Applications*. Springer, Dordrecht (2010). https://doi.org/10.1007/978-90-481-8847-5_5
14. Ning, H., Shihan, D.: Structure-based ontology evaluation. In: 2006 IEEE International Conference on e-Business Engineering, ICEBE 2006, pp. 132–137. IEEE, October 2006
15. Brank, J., Grobelnik, M., Mladenic, D.: A survey of ontology evaluation techniques. In: Conference on Data Mining and Data Warehouses, SiKDD 2005, Ljubljana, Slovenia, p. 4 (2005)
16. Arpinar, I.B., Giriloganathan, K., Aleman-Meza, B.: Ontology quality by detection of conflicts in metadata. In: Proceedings of the 4th International EON Workshop, May 2006
17. Bandeira, J., et al.: FOCA: a methodology for ontology evaluation, vol. 3, pp. 1–3 (2016). <http://arxiv.org/abs/1612.03353>
18. Maiga, G., Ddembe, W.: A flexible approach for user evaluation of biomedical ontologies (2008)
19. Brewster, C., et al.: Data driven ontology evaluation. In: 4th International Conference on Language Resources and Evaluation, LREC 2004, Lisbon, Portugal, p. 4 (2004)
20. Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J.: Modelling Ontology Evaluation and Validation. In: Sure, Y., Domingue, J. (eds.) *ESWC 2006*. LNCS, vol. 4011, pp. 140–154. Springer, Heidelberg (2006). https://doi.org/10.1007/11762256_13
21. Obrst, L., Ceusters, W., Mani, I., Ray, S., Smith, B.: The evaluation of ontologies. In: Baker, C.J.O., Cheung, K.H. (eds.) *Semantic Web*, pp. 139–158. Springer, Boston (2007). https://doi.org/10.1007/978-0-387-48438-9_8

22. Yu, J., Thom, J.A., Tam, A.: Ontology evaluation using wikipedia categories for browsing. In: Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, pp. 223–232. ACM, November 2007
23. McDaniel, M., Storey, V.C., Sugumaran, V.: The role of community acceptance in assessing ontology quality. In: Métais, E., Meziane, F., Saraee, M., Sugumaran, V., Vadera, S. (eds.) NLDB 2016. LNCS, vol. 9612, pp. 24–36. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41754-7_3
24. Supekar, K.: A peer-review approach for ontology evaluation. In: 8th International Protege Conference, pp. 77–79, July 2005
25. Maedche, A., Staab, S.: Measuring similarity between ontologies. In: Gómez-Pérez, A., Benjamins, V.R. (eds.) EKAW 2002. LNCS (LNAI), vol. 2473, pp. 251–263. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45810-7_24
26. Fahad, M., Qadir, M.A.: A framework for ontology evaluation. ICCS Suppl. **354**, 149–158 (2008)
27. Netzer, Y., Gabay, D., Adler, M., Goldberg, Y., Elhadad, M.: Ontology evaluation through text classification. In: Chen, L., et al. (eds.) APWeb/WAIM -2009. LNCS, vol. 5731, pp. 210–221. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03996-6_20
28. Martínez-Romero, M., Jonquet, C., O’connor, M.J., Graybeal, J., Pazos, A., Musen, M.A.: NCBO ontology recommender 2.0: an enhanced approach for biomedical ontology recommendation, *J. Biomed. Semant.*, **8**(1), 21 (2017)
29. Porzel, R., Malaka, R.: A task-based approach for ontology evaluation. In: ECAI Workshop on Ontology Learning and Population, Valencia, Spain, pp. 1–6, August 2004
30. Kehagias, D.D., Papadimitriou, I., Hois, J., Tzovaras, D., Bateman, J.: A methodological approach for ontology evaluation and refinement. In: ASK-IT Final Conference. June (Cit. on p.), pp. 1–13, June 2008
31. Burton-Jones, A., Storey, V.C., Sugumaran, V., Ahluwalia, P.: A semiotic metrics suite for assessing the quality of ontologies. *Data Knowl. Eng.* **55**(1), 84–102 (2005)
32. Buitelaar, P., Eigner, T., Declerck, T.: OntoSelect: a dynamic ontology library with support for ontology selection. In: Proceedings of the Demo Session at the International Semantic Web Conference (2004)
33. Buitelaar, P., Eigner, T.: Ontology search with the ontoselect ontology library. In: LREC, May 2008
34. Talebpour, M., Sykora, M., Jackson, T.W.: The role of community and social metrics in ontology evaluation: an interview study of ontology reuse. In: 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, pp. 119–127 (2017) <https://doi.org/10.5220/0006589201190127>
35. Fernández, M., Overbeeke, C., Sabou, M., Motta, E.: What makes a good ontology? A case-study in fine-grained knowledge reuse. In: Gómez-Pérez, A., Yu, Y., Ding, Y. (eds.) ASWC 2009. LNCS, vol. 5926, pp. 61–75. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-10871-6_5
36. Gill, T.: *Metadata and the Web: Introduction to metadata*, vol. 3, pp. 20–38. Getty publications, Los Angeles (2008)
37. Sowa, J.F.: Ontology, metadata, and semiotics. In: Ganter, B., Mineau, G.W. (eds.) ICCS-ConceptStruct 2000. LNCS (LNAI), vol. 1867, pp. 55–81. Springer, Heidelberg (2000). https://doi.org/10.1007/10722280_5
38. Lozano-Tello, A., Gómez-Pérez, A.: Ontometric: a method to choose the appropriate ontology. *J. Database Manage. (JDM)* **15**(2), 1–18 (2004)
39. Ding, L.: Swoogle: a search and metadata engine for the semantic web. In: Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management, pp. 652–659. ACM, November 2004

40. Hartmann, J., Sure, Y., Haase, P., Palma, R., Suarez-Figueroa, M.: OMV—ontology metadata vocabulary. In: ISWC, vol. 3729, November 2005
41. Matentzoglou, N., Malone, J., Mungall, C., Stevens, R.: MIRO: guidelines for minimum information for the reporting of an ontology. *J. Biomed. Semant.* **9**(1), 6 (2018)
42. Lewen, H., d’Aquin, M.: Extending open rating systems for ontology ranking and reuse. In: Cimiano, P., Pinto, H.S. (eds.) EKAW 2010. LNCS (LNAI), vol. 6317, pp. 441–450. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16438-5_34
43. Lewen, H., Supekar, K., Noy, N.F., Musen, M.A.: Topic-specific trust and open rating systems: an approach for ontology evaluation. In: Workshop on Evaluation of Ontologies for the Web, May 2006
44. Morse, J.M.: *Mixed Method Design: Principles and Procedures*. Routledge, London (2016)
45. Talebpour, M., Sykora, M., Jackson, T.: Social and community related themes in ontology evaluation: findings from an interview study. In: Fred, A., et al. (eds.) IC3K 2017. CCIS, vol. 976, pp. 320–336. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15640-4_16
46. Supekar, K., Patel, C., Lee, Y., Characterizing Quality of Knowledge on Semantic Web. In: FLAIRS Conference, pp. 472–478, May 2004
47. Wang, X., Guo, L., Fang, J.: Automated ontology selection based on description logic. In: Proceedings 2008 12th International Conference on Computer Supported Cooperative Work in Design, CSCWD, vol. 1, pp. 482–487 (2008)
48. Martínez-Romero, M., Vázquez-Naya, J.M., Pereira, J., Pazos, A.: BiOSS: a system for biomedical ontology selection. *Comput. Methods Programs Biomed.* **114**(1), 125–140 (2014)