Social and Community Related Themes in Ontology Evaluation: Findings from an Interview Study

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 deep exploration of what the term "quality" implicates in the field of ontology selection and reuse takes us much further than what the literature has mostly focused on, that is the internal characteristics of ontologies. A qualitative study with interviews of ontologists and knowledge engineers in different domains, ranging from biomedical field to manufacturing industry reveals novel social and community related themes, that have long been neglected. These themes include responsiveness of the developer team or organization, knowing and trusting the developer team, regular updates and maintenance, and many others. This paper explores such connections, arguing that community and social aspects of ontologies are generally linked to their quality. We believe that this work represents a significant contribution to the field of ontology evaluation, with the hope that the research community can further draw on these initial findings in developing relevant social quality metrics for ontology evaluation and selection.

Keywords: Ontology quality · Social quality metrics · Ontology reuse

1 Introduction

Ontologies play a major role in the field of knowledge and information management by furnishing the semantics to heterogeneous systems and [1], bridging multiple domains and enabling linked data [2]. They are also employed in different domains for various purposes. Ontologies provide many benefits, no matter in which domain they are used. They not only facilitate communication and knowledge transfer between systems, between humans, and between humans and systems [3], by uniquely identifying the meaning of different concepts in any domain, but they can also avoid the costs associated with new developments of knowledge models.

Despite the significant role that ontologies play in the semantic web, there is still little understanding about the way they should be built and developed [4]. Some believe that the cost of building and maintaining ontologies in certain domains can outweigh the potential benefits gained by using them [2, 5]. To deal with this concern, some have suggested reusing previously built ontologies, since this will help in achieving one of the main goals of ontology construction, that is to share and reuse

semantics [6], and will also save significant amount of time and financial resources. [7] believe that the future of construction of large-scale knowledge-based systems is highly dependent on reusing the components built by others, while such issues are still at the forefront of ontology engineering [8].

Regardless of all the advantages of reusing ontologies and the availability of different ontologies, ontology reuse has always been a challenging task [7]. Methods for building ontologies are usually blamed for lack of reuse strategies and some argue that these methodologies are not explicitly concerned with ontology reuse [9]. 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 [10], as the hardest step in the process of ontology reuse. Researchers not only have to find the most appropriate ontologies for their search query, but they should also be able to evaluate those ontologies according to different implicit or explicit criteria.

This study aims to address some of the challenges faced in the first steps of the general process of reusing ontologies, which is to evaluate and then select the right ontology for reuse. This study contributes with qualitative data and findings to this ongoing challenge by documenting the process of selecting an ontology for reuse. It differs from previous studies, which focused purely on evaluating different pre-selected metrics. In this study, our focus was to qualitatively understand the process and reasoning behind ontology selection and reuse, with a particular focus on the under-researched social and community aspects of ontology quality. Interviews were used to understand how ontologists and knowledge engineers in different domains search for, evaluate and select an ontology for reuse. This paper is an extended version of [11] and aims to study and explore (1) the main characteristics of a reusable ontology, (2) the main metrics used to evaluate the quality of an ontology before selecting it for reuse, and (3) the link between community related metrics e.g. who has built the ontology, who has used the ontology, etc. and ontology evaluation for selection and reuse.

2   Background

Ontology evaluation is one of the most popular topics 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, measuring its fitness for a specific purpose, etc. There are many different ways of defining ontology evaluation; one of the most popular and also the earliest definitions was provided by [12] where the term evaluation was used to refer to the technical judgment of an ontology, considering different aspects of it, such as the definitions of its components, documentation and software environment. 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 concerned with determining how well an ontology corresponds to what it should represent [13]. In the other word, ontology validation focuses on building the correct ontology whereas ontology verification is about building an ontology correctly [14].

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 [15]. This definition is used to link the process of ontology evaluation to ontology selection. The aim of ontology selection is to identify an ontology, an ontology module or a set of ontologies that satisfy a certain set of criteria or selection requirements [16]. 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, the libraries that the selection is based on, etc. [15]. The term "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 like usability, usefulness, etc. [17]. Unlike the first definition of the ontology evaluation [12], 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.

Ontology evaluation can also refer to a function or an activity that aims to map an ontology or a component of an ontology e.g. its concepts to a score or a number, e.g. usually in range of 0 to 1 [18]. The main aim of these types of processes/functions is to measure and assess the quality of an ontology with regards to a set of predefined metrics or requirements [19]. This definition is somehow similar to what [17] define as ontology quality assurance that 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 according to 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 a descending order of quality, according to the score that is assigned to them in the evaluation process.

There are many different reasons why ontologies are evaluated. The main goals can be categorized as either for quality assessment, for tracking progress in ontology evolution, or for ranking, each of these are briefly described next.

Evaluation for correctness; refers to those approaches that aim to measure the logical and formal correctness of an ontology content [20, 21]. One of the most well-known applications for evaluating the correctness of ontologies is OntOlogy Pitfall Scanner (OOPS) [22]. OOPS is based on a catalogue of 40 bad practices or pitfalls that can happen in ontology development and it is able to automatically detect them [22]. ODEval is also developed by [23] to evaluate knowledge representation of RDF(S), OWL, and DAML + OIL concept taxonomies by capturing different circularity and redundancy problems as well as partition errors.

Evaluation for quality assessment; assessing the quality of ontologies is another main reason for evaluating them. This type of evaluation is mostly done by end users and helps in the process of ontology selection for reuse. Ontology assessment is a very challenging task both because of lack of generic quality evaluation solution and metric and also because determining the right elements of quality to evaluate is difficult [24].

Evaluation for tracking progress in ontology evolution [19, 25]; the aim of this type of evaluation is to track different characteristics and changes of ontologies over time and over different versions [26]. ONTO-EVOAL approach, for example, uses patter

modelling to make sure that consistency and quality is maintained in the process of ontology evolution [27].

Evaluation for Ranking [21]; search and selection systems in the field of ontology engineering usually show a ranked list of ontologies in output. To do that, they should first be able to evaluate ontologies according to different query-dependent and query independent criteria, then assign some score to the ontologies based on how well they are covering or meeting those criteria, and finally rank ontologies based on their score. In Swoogle, for example, a PageRank [28] like algorithm is used to calculate the rank of each ontology based on the number of link from and to those ontologies [29].

Ontology evaluation is important in the ontology development process, whether it is built from scratch, automatically or by reusing other ontologies [30]. While building an ontology from scratch, developers need to evaluate the outcome ontology, to measure its quality [31], to check if it meets their application requirements [30] and also to identify the potential refinement steps [32]. Evaluation is also helpful in checking the homogeneity and consistency of an ontology when it is automatically populated from different resources [26, 30]. Building an ontology from scratch is a very costly as well as a time-consuming task [33, 34]; therefore, ontologists are urged to consider reusing exiting ontologies before building a new one [35]. Ontology evaluation is and has always been a major concept when it comes ontology reuse [36]. 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 [32, 33, 36, 37].

Moreover, the number of ontologies on the web has been increasing rapidly [30] and users usually face multiple ontologies when they need to choose or use one in their everyday activities [19, 32, 38]. Before using an ontology in an application or selecting it for reuse, ontologists have to assess its quality and correctness and also compare it with 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 [16] and is used to select the best or the most appropriate ontology among many other candidates in a domain [32]. Evaluating an ontology is considered as a complicated process [19, 39]; it is believed that failure to evaluate ontologies and to choose the right ontology might lead to using the ones with a lower quality [19].

There are various ontology evaluation methods and several ways of classifying them in the literature. According to [32], ontology evaluation can be done in four major ways: (1) evaluating an ontology by comparing it to a "golden standard", (2) evaluating an ontology by comparing it to a source of data, (3) evaluating an ontology by running it in an application as part of a system and evaluating the resulting performance, and (4) asking human experts to evaluate an ontology against a set of predefined quality criteria.

Beside the above-mentioned methods, that are very popular in the literature, there are some other ways of classifying ontology evaluation approaches. For example, they can be classified based on the type of the metrics they use to assess ontologies. 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 [40]. Qualitative approaches can also be used to evaluate an ontology based on the principles that are/were used in its construction [35]. Other evaluation approaches in the literature are based on different quantitative criteria about different aspects of ontologies such as its structure,

content, etc. These approaches, that are also known as formal rational approaches, are usually concerned with technical and economic aspects of ontologies and use different goal based strategies [33].

Moreover, ontology evaluation approaches can be glass-box or black-box. Glass-box approaches tend to evaluate the internal content and structure of ontologies [36]; they 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 [36]. Ontologies can also be evaluated as a whole or according to their different layers e.g. data level, taxonomy level, application level, etc. [34] has divided the concept of ontology quality to two broad types: "Total Quality" and "Partial Quality". Some believe that evaluating an ontology as a whole, specially automatically, is not possible or practical, especially considering the complex structure of ontologies [32].

From all the methods, metric-based approaches (4) are very popular and different researchers have attempted to introduce various metrics and measures that can be used to evaluate ontologies and help in the decision-making process for ontology selection. The aim of this method, that is also called featured-based approach, is to offer a quantitative perspective of evaluating ontologies by gathering data and meta-data on different aspect of the ontology [26].

Ontometric [41], as one of the most popular examples of this approach, consists of a detailed set of 160 criteria to examine different dimensions of ontologies namely content, language, ontology construction methodologies, costs, and tools. While many of the criteria in metric-based evaluation approaches aim to measure different internal components of an ontology e.g. structure, content, coverage, etc., some of these have focused on non-ontological and social aspects [39] of ontologies like popularity [42–44].

Besides how ontologies are built and what they are covering or even not covering, how they are used by communities is one of the dimensions and important aspects of ontologies that can be used to evaluate or select them. [14] define user-based ontology evaluation as the process of evaluating an ontology though users' experiences and by capturing different subjective information about ontologies. The term "Social Quality" has also been used to reflect the existence of ontologies as well as agents and users in communities [24] and refers to the level of agreement about an ontology and among different participants or members of a community [39]. [39] argue that there is a link between the quality of an ontology and community approval and participation in its evolution.

Popularity or acceptance are two of the most popular and used terms in the literature to refer to the social aspects of ontologies' quality; however, there is still no consensus on the definition of these terms. Popularity and acceptance tend to be mostly used to refer to the number of times an ontology has been viewed or used in a specific repository. NCBO Ontology Recommender [45] for example, calculates the popularity of an ontology by checking the presence of the ontology in well-known repositories as well as looking into the number of visits or pageviews to an ontology in ontology repositories in a recent specific period [45]. In the paper by [24] the authors also refer to the term history to indicate the number of times an ontology has been used.

The second definition of popularity is based on applying the PageRank algorithm [28] to ontology engineering field and focuses on the import feature of ontologies. [43] for example has defined the term "direct popularity" as the number of ontologies importing a given ontology. [44] used the same definition to define what they call popularity, that for them is measured by considering how much an ontology is referenced by others. As a part of the authority metric in [46], authors have mentioned a metric called citation and have defined it as the number of occurrence of daml:sameClassAs, rdfs:seeAlso, owl:imports in a given ontology.

Social quality plays an important role in ontology evaluation for selection and reuse. According to a study that was conducted by [47], 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. [39] also highlight the importance of relying on the wisdom of crowd in ontology evaluation and believe that improving the overall quality of ontological content on the web is a shared responsibility within a community.

3   Methodology

Semi-structured interviews with ontologists and knowledge engineers were conducted to investigate the thinking behind and the processes commonly involved in evaluating ontologies for their reuse. Purposive sampling was used to find the experts in the field of ontology engineering [48]. Different sampling strategy namely intensity sampling was applied to find the ontologies that have been reused and then to interview the individuals who had built or had reused those ontologies [49]. Moreover, homogenous sampling was used to find different ontology related research groups in different organizations and universities working in different domains.

Table 1. Domain expertise of ontologists and knowledge engineers interviewed.

Code	Domain, organization, or project
NBI1	Ontologist/IBM, Smarter Planet Project
NBI2	Professor, Manufacturing Informatics
NBI3	Ontology engineer/Semantic Web
NBI4	Researcher/Laboratory for Applied Ontology
NBI5	Researcher/Smart Cities, Geo Ontologies
SB1	Ontology developer/Industry, W3C, NHS
SB2	Researcher/BioPortal
SB3	CEO and ontology developer/Bioinformatics
SB4	Lecturer/Computing Science and Biology
SB5	Research scientist/Protégé group
NBI6	Researcher/Industrial ontologies
BI1	Group leader/Bioinformatics, Gene ontology
BI2	Researcher/BioPortal
BI3	Ontology Developer/Bioinformatics, Gene ontology
BI4	Researcher/Biomedical Informatics

The suitability of the interview questions was tested in a pilot phase to establish the relevance of the interview questions to the research. Five ontologists and knowledge experts took part in this first pilot round of interviews. All these interviews were recorded, with permission, and were transcribed and thematically analyzed. The pilot phase was also helpful in determining the time it takes to conduct the interview as well as the flow and order of the interview questions.

15 researchers with different levels of expertise and knowledge engineering backgrounds were interviewed. As it is seen in the Table 1, four out of the fifteen interviewees had only worked in the biomedical field, five had some biomedical experience but had also worked in other fields such as computer science, and the rest of the interviewees were mostly involved in manufacturing, smart cities, etc. The semi-structured interview protocol focused on how each individual (i) built, (ii) searched for, (iii) evaluated and (iv) reused ontologies. Interviews ranged from 20 to 60 min, all of which were conducted via Skype. Interviews were recorded, and the interviewer took field notes during the interview. Field notes and transcriptions were coded using NVivo.

Interviews were conducted until no new information or theme was found [50] and the conceptual saturation was reached. The sample size can also be justified by some of the previous similar research on ontology evaluation for example the survey that was conducted by [51], which had 10 participants. Based upon the research questions, we began by coding for the following themes:

1. Building a reusable ontology
2. Characteristics of a reusable ontology
3. Finding a reusable ontology
4. Evaluating/trusting/selecting ontologies
5. The importance of community.

4   Findings

According to the interview findings, metrics for evaluating the quality of an ontology for reuse can be classified into the following categories:

- Metrics based on different internal components of ontologies including content, structure, coverage, etc.
- Metrics related to different metadata about an ontology such as methodology used, availability of documentation, language, etc.
- Metrics based on the social aspects of ontologies like community, popularity, ontology developer team, etc.

The following parts of this paper moves on to describe in detail different metrics that were considered and used by the participants of this study in the evaluation process as well as how they referred to the community to search for, find and evaluate an ontology for reuse.

4.1   Evaluation Based on Internal Aspects of Ontologies

As it is seen in the literature, many of the evaluation methods are based on different internal components of ontologies such as content, structure, etc. When asked about the quality of ontologies and ontology evaluation, some of the interviewees made a link between ontology content and definitions and their quality. SB3 and SB4 for example pointed out that they check the correctness of an ontology's content before selecting it for reuse. Ontologists and knowledge engineers not only care about the content of an ontology, but they consider other content related metrics such as consistency, correctness, coverage, etc. NBI1 for example mentioned coverage and the relationship between concepts as two of the metrics that he uses for ontology evaluation. Some of the interviewees mentioned being "well-structured" as one of the characteristics of a good, reusable ontology. According to them, a well-structured ontology can be defined as an ontology that has some rich (NBI4) and correct (SB1) connections between its categories. Ontology syntax, clear definitions and scope were among the other internal characteristics of ontologies that can be used for evaluating them.

4.2   Evaluation Based on Metadata About Ontologies

Besides the internal components of ontologies, wide range of metadata and additional elements of information about different aspects of ontologies can also be used for their evaluation. Many of the respondents found having access to additional information on ontologies, both in form of labels and comments for/on different ontology components or as external documentation, to be very helpful. Moreover, some of the interviewees mentioned that before selecting an ontology for reuse, they would like to know if the ontology is based on any standard, methodology or a common framework. Respondents also emphasized the importance of reusing other ontologies while building a new one and said that before selecting an ontology for reuse, they will check and see if the ontology has reused other ontologies and if it is based on upper level ontologies. Language that an ontology is built in and its size were among the other metadata about ontologies that can be used for their evaluation.

4.3   Evaluation Based on Community and Social Aspects

Communities can affect different aspects of the process of ontology selection for reuse, from how users search for and find reusable ontologies to how they evaluate those ontologies before selecting them for reuse. This section will first discuss how community and social aspects of ontologies help in the ontology search and discovery process and will then move on in exploring how different social related metrics can be used in the evaluation process.

Community and Ontology Search. One of the fundamental objectives of the interviews was to explore how ontologists and knowledge engineers search for reusable ontologies? Consequently, the question "how do you find the ontology you want to reuse?" was asked and while the researcher was expecting to hear about some of the popular search engines in ontology engineering domain like Swoogle, BioPortal, etc.,

literature and published papers were mentioned by many of the interviewees as one of their main sources of finding the ontologies they need.

Interviewee NBI4 for example, blamed his domain for lack of good and well-established repositories for ontologies and said that "I go to the literature". Another interviewee, SB3, also emphasized the significant role of literature in the process of searching for ontologies and mentioned that "reading publications around the ontology" is a very good method to help find the ontology, especially if someone is new to the field.

Besides helping to find a reusable ontology, some of the other interviewees stated that they use the literature and research papers as a tool to evaluate the quality of an ontology. Respondent NBI4 pointed out:

"If an ontology is good and is used, you find a cite in the literature."

Being based on published research papers will not only affect the quality of an ontology, but according to some of the respondents, will also affect the popularity of an ontology; BI4 for example stated:

"Popular ontologies are better ontologies, people are just familiar with popular ontologies so whenever you go to any ontology related conference, you will always have a workshop or a paper that talks about the ontology."

Community and Ontology Evaluation. As was highlighted in Sect. 2, various work has looked at the quality and evaluation of ontologies, however while some of the papers have attempted to cover the social aspects of ontology evaluation, none have gone further than measuring popularity, authority, and history of ontologies and almost all of them have neglected the other interactions in the community that can affect the way ontologies are evaluated, selected, and reused. Hence to explore the role of community in ontology sharing and reuse, participants were asked how interactions with people in their domain may affect the way they tend to evaluate an ontology for reuse. According to the interviews, participants not only use the community to evaluate an ontology before selecting it for reuse, but some of them also evaluate the ontologies they are building by the feedback they receive from the community.

Build Related Information. Several researchers mentioned the importance of different types of build related information such as who/which organization has built the ontology, what the ontology has been built for e.g. the use case, who are the different stakeholders of the ontology, how the ontology was built (e.g. in collaboration), etc. Interestingly, one of the first things interviewees would say was that to evaluate an ontology, they will ask themselves if they know the developer of the ontology?

Interviewee BI3 for example emphasized the importance of knowing the developer team and its effect on the reuse process:

"I have to say, in reusing thing, there is often politics and connections are as important as anything else. So, it is not always the best one that wins."

He also added, quality of an ontology may sometimes come second:

"You know there might be constraint in terms of I may not like a particular ontology but because a bunch of other people are using it and I want to standardize with them, I might use it anyway."

Respondent SB4 also brought up the issue of trusting the developer team:

"Science is a social enterprise, I mean this is how everything works in science, you know if you look at a paper, do you trust the paper? you look at the authors first and then you read the paper and then you pick about what they have done but yes I mean it is a major criteria, major quality criteria, it may or may not right; it is a bit of old boys club but yes that is how people make decision. I normally read the definitions and then go to other things; do I trust the people who are making it?"

Besides the information about the developer team or organization, some of the respondents would consider the reasons that ontology was built and used for before selecting an ontology. They were also interested in having some information about the stakeholders of the ontologies. Interviewee SB3 said:

"Completely separated from the people developing it, are there other people who uses this ontology beyond just that group, that tells you something about it. I think also finding out how they are using it, is also important, you know what data is being annotated with those ontology is also important question, but I have some data and I know I want to integrate with something done in another institute, what is the ontology there they are using, that is also important, so I think there is a list of the things you want to check!"

Regular Updates and Maintenance. Ontology maintainability is one of the significant metrics while evaluating the quality of an ontology and before selecting it for reuse. In the interviews, there were numerous examples of linking the quality of an ontology to how regularly it is updated and maintained. For some participants like NBI3, regularity of updates was the first thing that they would look at when evaluating a particular ontology:

"Somebody build ontology during his research in 1998 and he stored it on the web and then he left it, it is available but not updated, things will get obsolete very soon so we make sure to use the ontologies which are regularly updated, it is the first thing."

Some of the respondents like SB3 compared maintenance with some of the very popular quality metrics in the literature like coverage and said:

"Does it have my terms? I think is important but there are many others that you need to consider when you are picking an ontology beyond just does it have the words in ontology, about maintenance, do they update regularly, do they release regularly? do they have a record of doing that? How responsive they are to updates when you need new terms? all that sort of stuff. If they are publishing it once every two years it is probably not a good ontology."

Other participants like BI1 firmly believed that updates and maintenance play a very important role in their domain and said:

"No way that an ontology is keeping on in biology not getting updated, biology is changing too fast so all the relevant ontologies in biology are getting updated."

Interviewee NBI2 also made a link between the nature of the domain that he is working in and the necessity of regular updates:

"It is about flexibility, if you want to, in manufacturing business things are changing all the time, so you need solutions that are easy and flexible to stay in, to stay relevant to what you are doing tomorrow as well as what you are trying to do today."

Interviewee BI3 compared the ontology engineering with software engineering and said:

"If you are going to reuse a piece of open source software you will do the same thing, you will open the GitHub website and say you know if you looked in it and nobody updated it or anything in three years, you might think no; whereas if it looks like there is an active ongoing community, you will think yes, if I have problems I can ask people and I can get bugs fixed."

BI4 believed that there is a link between the popularity of an ontology and the regularity of updating it and said:

"It might be useful to use popular ones because there are the ones that are mostly updated so gene ontology has a release I think every day or every 12 h, so the popular ontologies are the ones that are most updated."

Not only the regularity of updates is important, but also how people deal with it is the other important issue. Respondent SB3 talked about the importance of having an update mechanism and said:

"I think in the field that I am working, there are other challenges, one of which is how you deal with update mechanism of ontologies, if you annotate data to ontology which is typically use case for how you keep up-to-date with the fact that ontologies change reasonably often, you might have a big database of data, that you used the data in, new ontologies come along, the effect the way the data has been represented in your database, gotta have a update mechanisms for dealing with that and that can be tricky actually, it is not as simple often as swapping things out when something gets made obsolete, it is replaced with other things, you have to deal with."

Responsiveness. Responsiveness of the ontology developer team was among one of the other widely mentioned criteria when evaluating the quality of ontologies for reuse. Some of the respondents argued that not only knowing the developer team or organization is important, but also having an active ongoing community and their willingness to collaborate, evolve and develop the ontology further is an important factor when assessing an ontology. Interviewee BI3 put it in this way when he was asked about the importance of responsiveness:

"I would say it is definitely high up; I mean having someone at the other end of line that you feel that you can trust is definitely very important. If it looks like there is an active ongoing community, you will think yes if I have problems I can ask people and I can get bugs fixed."

SB5 used one of the popular ontologies in her field as an example and added:

"For example, the fact that the Gene ontology has a huge community behind it is important because it means that they have a curation process in place and quality assurance and so on; so that kind of gives more confidence that the ontology is as good as it can be, it is not perfect for sure but I mean that it is vetted by the community."

Respondent BI1 chose responsiveness as the first quality metric he would consider for evaluating an ontology and compared it with one of the very popular ontology evaluation metrics, that is availability of documentation:

"I would say the responsive of the team obviously is the top-quality metric for me, because nothing is perfect but if something gets improved then it will get good like if you have a question, you need to add a term, something does not make sense, you contact them, they answer and they answer in a constructive way; this is good because all the ontologies are work in progress, there is no finished ontology in my domain."

Popularity. When asked about the link between popularity of an ontology and its quality, participants had interesting thoughts and responses. Most of the interviewees defined popularity as the number of times an ontology has been viewed or used in a repository. However, some of the interviewees doubted the importance of this metric. The responses can be classified into the following groups: (1) those who were against this metric, (2) those who liked popularity as a quality metric for ontology evaluation but did not agree with the way it was being computed and (3) those who found this metric useful.

The first group of respondents believed that the popularity of an ontology or the number of times it has been used is not that important. As interviewee BI1 would put it:

"To me it would not be very important except if two ontologies are really very equal in everything else, I will take the most used ones, but I do not think, it is not really relevant to me, if it is the right tool for the job, it is the right tool!"

It was also believed that the number of times an ontology is used depends on different factors such as its size, level of specialization and the domain that it is built in and cannot be considered as a metric to measure its quality. According to interviewee BI1:

"Some ontologies are more specialized, so less people use them because it corresponds to a very special need, but may this people, are the right people and are using it well."

Interviewee SB3 also linked the use of an ontology to its size and added:

"If there is a small ontology but really focused on representing an area that has not been done before but it is correct, it is absolutely correct, I think that is perfectly reasonable, even if it is not widely used."

Some other interviewees like NBI5 found popularity a helpful metric, but believed that it is highly dependent on the domain that the ontology is used in:

"It depends on the domain that it has been reused in, if it is just medical domain, it is difficult to say that it is a reusable ontology!"

The second group agreed on the necessity of having such a metric to identify the more popular ontologies in different domains but were not sure about the usefulness of the current methods that are used to measure the popularity. As interviewee NBI3 would put it:

"How many times an ontology is viewed will not help you, I may click just for exploration, and I will say it is not my thing and I don't want it; it shows how catchy the term is or how important, how regularly, how often this term is chosen, but it does not mean the use of the ontology; so, I think there should be some other way."

BI4 used a very interesting personal experience to prove the inaccuracy of the current techniques of measuring the popularity:

"When we were visualizing all the user exploration on ontologies on BioPortal, and we found that gene ontology is not accessed that much using BioPortal and I thought that it was very surprising because the gene ontology is very famous and then I found out because there is a gene ontology browser called AmiGo, and their visualizer tool is much better than BioPortal visualization of gene ontology, so people generally go to gene ontology website and lunch the AmiGo browser and go to gene ontology there, so you can say that gene ontology is much more accepted but if you just look at the clicks (in BioPortal) and you might say that gene ontology is not that much famous."

Interviewee SB3 also thought that having a quality metric like popularity is a step in the right direction but believed that it might be misleading by causing a snowball effect; according to him:

"I can see that you can also putting a little metric for usage or browsing or how many people read these things, that is a kind of useful but it does not tell you the whole picture, you know you can end up with a false signal there; you recommended an ontology because it is useful because someone uses it and then you recommend it so someone else uses it and so on and so on, what I mean, so you are getting in that cycle of, it grows and grows!"

The last and also the minority group were those who thought it worth having a metric like popularity and highlighted the importance of community acceptance. According to interviewee NBI4:

"If a community is using the ontology and is happy with it, I take thing to account, so I try to reuse or to do something to extend it or maybe very careful on changing it. I need to have motivations because after all ontologies should have people working in the domain and so if they are happy with that one and I see things that are no good, I point it out and I may suggest an extension, whatever but I try to reuse what I have."

The other definition of this metric that focuses on the link between popularity and the number of imported ontologies was also brought up by some of the respondents. NBI5 for example, made a link between the quality of an ontology and the fact that the ontology has reused other ontologies and said:

"The quality of an ontology depends on the relation between the ontology to upper level ontologies; the more 'same-as', 'equivalent-as' links I can find in an ontology. It also can be seen as a sign or a feature of the ontology that can be reused because if it is 'same -as' a concept that we already know, then it can be replaced."

NBI6 also believed that reusing some of the ontologies are inevitable and not importing will seem as a negative impression:

"Whenever I have an ontology where there is a person, I will never ever create my own person class, I will always reuse FOAF. I think it would be ridiculous to create my own class and some of those are very very strong class definition, so it will always worth reusing and I think it will be even mistake by ontology engineer to develop their own class and for me, if I see an ontology doing that, I will get a negative impression."

5   Discussion

The notion of ontology quality and the process of evaluating it is one of the most significant and also complicated components in the field of ontology engineering. The challenge of choosing the right or best ontology for a task or reuse is what ontologists and knowledge engineers face on a daily basis. Despite the importance of this matter and the widespread research on this topic, there are still many unanswered questions and challenges in ontology evaluation. Interviewees have shown a range of different responses when they were asked about the ontology evaluation process; some of them mentioned the different well-known internal quality metrics such as ontology content, structure, and others; however, and surprisingly, the main focus of most of the responses was on different metadata and social aspects related to ontologies. The scope

of this study has particularly focused on the ways communities can help in the process of ontology selection and evaluation.

As it was discussed before, community related factors such as reputation of the developer team or organization in the domain and regularity of updates were highlighted as some very important metrics to be considered when evaluating and selecting an ontology for reuse. The results of this study have also found that the quality of ontologies is generally considered to be limited; ontologists and knowledge engineers have pointed out that either way there isn't such a thing as a complete or finished ontology, hence ontologists often need to count on the responsiveness of the ontology developer team and organization as well as their attitude toward the requests for changes. However, this has not previously been described and most of the existing studies have failed to cover and analyze the association between ontology quality and evaluation and the role of the community in that process.

Ontology popularity [42] is one of the most commonly defined and used social quality metric in the literature and many of the prior studies have noted the importance of this metric. However, some of the interviewees doubted the link between the quality of an ontology and its popularity. According to the interviews, respondents care more about the projects that the ontology has been or is being used in, compared to the number of times it was used. Regarding the second definition of popularity, that is more about the linkage and the citation between ontologies [46], it seems that further research should be undertaken to investigate the importance of this factor and the way it can be employed to calculate popularity of an ontology.

Overall, the evidence from this exploratory study suggests that there is a clear interest for community-based ontology evaluation and the need for relevant metrics. Further research is needed to confirm the quality metrics suggested in these research interviews and what their relative importance may be, whether there are differences in ontology engineering domains, or other important idiosyncrasies deserving further attention. To provide more generalizable findings for this research, the next stage of our research agenda will be to conduct large scale data collection via a survey targeting ontology engineers from heterogeneous domains. The expected outcome would be to introduce a community-based quality metrics as well as to design and implement suggestions and guidelines that will help in designing and implementing ontologies that can be more easily found and reused, based on community measures identified through this ongoing research work.

6   Conclusion

The main goal of the current study was to determine and explore the set of steps that ontologists and knowledge engineers tend to take in different phases of ontology selection process, from ontology search and discovery to ontology evaluation and selection. The study has found that while internal characteristics of ontologies like content, structure and consistency are considered in the evaluation process, ontology evaluation is mostly focused on non-ontological features of ontologies such as metadata about ontologies, social and community related quality metrics. It was also shown that ontologies are usually considered as incomplete ongoing projects; therefore, it is

important to know and also to trust the ontology developer and maintenance team or organization before selecting it for reuse. These findings enhance understanding of the metrics used in the evaluation process and it is hoped that they can be of help to the community in the ontology selection process. A natural progression of this work would be a framework of non-ontological quality metrics for ontology evaluation and selection.

References

1. Shadbolt, N., Berners-Lee, T., Hall, W.: The semantic web revisited. IEEE Intell. Syst. **21** (3), 96 (2006)
2. El Kadiri, S., Kiritsis, D.: Ontologies in the context of product lifecycle management: state of the art literature review. Int. J. Prod. Res. **53**(18), 5657–5668 (2015)
3. Bürger, T., Simperl, E.: Measuring the benefits of ontologies. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM 2008. LNCS, vol. 5333, pp. 584–594. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88875-8_82
4. Ding, Y., Foo, S.: Ontology research and development. Part 2 - a review of ontology mapping and evolving. J. Inf. Sci. **28**(5), 375–388 (2002)
5. Alani, H., Brewster, C., Shadbolt, N.: Ranking ontologies with AKTiveRank. In: Cruz, I., et al. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 1–15. Springer, Heidelberg (2006). https://doi.org/10.1007/11926078_1
6. Simperl, E.: Reusing ontologies on the semantic web: a feasibility study. Data Knowl. Eng. **68**(10), 905–925 (2009)
7. Uschold, M., et al.: Ontology reuse and application. Form. Ontol. Inf. Syst. **179**, 192 (1998)
8. Presutti, V., Lodi, G., Nuzzolese, A., Gangemi, A., Peroni, S., Asprino, L.: The role of ontology design patterns in linked data projects. In: Comyn-Wattiau, I., Tanaka, K., Song, I.-Y., Yamamoto, S., Saeki, M. (eds.) ER 2016. LNCS, vol. 9974, pp. 113–121. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46397-1_9
9. Annamalai, M., Sterling, L.: Guidelines for constructing reusable domain ontologies. In: 3rd Workshop on Ontologies in Agent Systems Sofitel, no. July, pp. 71–74 (2003)
10. Bontas, E.P., Mochol, M., Tolksdorf, R.: Case studies on ontology reuse. In: Proceedings of IKNOW05 International Conference on Knowledge Management, vol. 74, pp. 345–353 (2005)
11. Talebpour, M., Sykora, M.D., 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)
12. Gómez-Pérez, A.: Some ideas and examples to evaluate ontologies. In: CAIA, p. 299 (1995)
13. Gómez-Pérez, A.: Evaluation of taxonomic knowledge in ontologies and knowledge bases (1999)
14. Hlomani, H., Stacey, D.: Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: a survey. Semant. Web J. **1**, 1–11 (2014)
15. Fernández, M., Cantador, I., Castells, P.: CORE: a tool for collaborative ontology reuse and evaluation. In: CEUR Workshop Proceedings, vol. 179 (2006)
16. Sabou, M., Lopez, V., Motta, E., Uren, V.: Ontology selection: ontology evaluation on the real semantic web. In: 4th International EON Workshop, Evaluation of Ontologies for the Web, EON 2006, vol. 179 (2006)

17. Suarez-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, pp. 1–15 (2008)
18. Brank, J., Mladenic, D., Grobelnik, M.: Gold standard based ontology evaluation using instance assignment. In: Proceedings of EON 2006 Workshop (2006)
19. Yu, J., Thom, J.A., Tam, A.: Requirements-oriented methodology for evaluating ontologies. Inf. Syst. **34**(8), 686–711 (2009)
20. Jonquet, C., Musen, M.A., Shah, N.H.: Building a biomedical ontology recommender web service. Biomed. Seman. **1**, S1 (2010)
21. Duque-Ramos, A., et al.: Evaluation of the OQuaRE framework for ontology quality. Expert Syst. Appl. **40**(7), 2696–2703 (2013)
22. Poveda-Villalón, M., Gómez-Pérez, A., Suárez-Figueroa, M.C.: Oops!(ontology pitfall scanner!): an on-line tool for ontology evaluation. Semant. Web Inf. Syst. **10**(2), 7–24 (2014)
23. Corcho, Ó., Gómez-Pérez, A., González-Cabero, R., Suárez-Figueroa, M.C.: ODEval: a tool for evaluating RDF(S), DAML+OIL, and OWL concept taxonomies. In: Bramer, M., Devedzic, V. (eds.) AIAI 2004. IIFIP, vol. 154, pp. 369–382. Springer, Boston, MA (2004). https://doi.org/10.1007/1-4020-8151-0_32
24. 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)
25. Yang, Z., Zhang, D., Ye, C.: Evaluation metrics for ontology complexity and evolution analysis. In: Proceedings - IEEE International Conference on E-Business Engineering, ICEBE 2006, no. 90204010, pp. 162–169 (2006)
26. Arpinar, I.B., Giriloganathan, K, Aleman-Meza, B.: Ontology quality by detection of conflicts in metadata. In: CEUR Workshop Proceedings, vol. 179 (2006)
27. Djedidi, R., Aufaure, M.A.: ONTO-EVOAL an ontology evolution approach guided by pattern modeling and quality evaluation. In: Proceedings of 6th International Symposium, Sofia, Bulgaria, 15–19 February 2010, FoIKS 2010, vol. 6 (2010)
28. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking: bringing order to the web. World Wide Web Internet Web Inf. Syst. **54**(1999-66), 1–17 (1998)
29. Ding, L., Finin, T., Reddivari, P., Cost, R.S., Sachs, J.: Swoogle : a search and metadata engine for the semantic web. In: ACM Conference on Information and Knowledge Management, pp. 652–659 (2004)
30. 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, pp. 115–130. Springer, Dordrecht (2010). https://doi.org/10.1007/978-90-481-8847-5_5
31. Ning, H.N.H., Shihan, D.S.D.: Structure-based ontology evaluation. In: 2006 EEE International Conference on E-Business Engineering, ICEBE 2006, pp. 2–7 (2006)
32. Brank, J., Grobelnik, M., Mladenic, D.: A survey of ontology evaluation techniques. In: Conference on Data Mining and Data Warehouses, SiKDD 2005, p. 4 (2005)
33. Maiga, G., Ddembe, W.: A flexible biomedical ontology selection tool. In: Kizza, M., Lynch, K., Nath, R., Aisbett, J., Vir, P. (eds.) Strengthening the Role of ICT in Development, vol. 5, pp. 171–189. Fountain Publishers (2009)
34. Bandeira, J., Bittencourt, I.I., Espinheira, P., Isotani, S.: FOCA: a methodology for ontology evaluation, vol. 3, pp. 1–3 (2016)
35. Brewster, C., Alani, H., Dasmahapatra, S., Wilks, Y.: Data driven ontology evaluation. In: 4th International Conference on Language Resources and Evaluation, LREC 2004, p. 4 (2004)
36. Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J.: Qood grid: a metaontology-based framework for ontology evaluation and selection. In: Proceedings of EON 2006, vol. 4011, pp. 140–154 (2006)

37. Obrst, L., Ceusters, W., Mani, I., Ray, S., Smith, B.: The evaluation of ontologies. Seman. Web, 139–158 (2007)
38. 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, CIKM 2007, p. 223 (2007)
39. 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
40. Porzel, R., Malaka, R.: A Task-based approach for ontology evaluation. Biomed. Semant. (2004)
41. Lozano-Tello, A., Gomez-Perez, A.: ONTOMETRIC: a method to choose the appropriate ontology. Database Manag. **15**(2), 1–18 (2004)
42. 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)
43. 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
44. Wang, X., Guo, L., Fang, J.: Automated ontology selection based on description logic. In: Proceedings of 2008 12th International Conference on Computer Supported Cooperative Work in Design, CSCWD, vol. 1, pp. 482–487 (2008)
45. 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. Seman. **8**(1), 21 (2017)
46. Supekar, K., Patel, C., Lee, Y.: Characterizing quality of knowledge on semantic web. In: 7th International Florida Artificial Intelligence Research Society Conference, pp. 220–228 (2004)
47. 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
48. Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N., Hoagwood, K.: Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm. Policy Ment. Health **42**(5), 2–4 (2015)
49. Suri, H.: Purposeful sampling in qualitative research synthesis. Qual. Res. J. **11**(2), 63–75 (2011)
50. Guest, G., Bunce, A., Johnson, L.: How many interviews are enough? An experiment with data saturation and variability. Field Methods **18**(1), 59–82 (2006)
51. Tello, A.J.: Métrica de idoneidad de ontologías. Universidad de Extremadura (2002)