Towards Personalized Support for Ontology Selection

Marwa Abdelreheim[1], Friederike Klan[2], Taysir Soliman[1]

[1] Department of Information Systems, Faculty of Computer and Information Sciences, Assiut University, Egypt
{marwa.abdelrehem, taysser.soliman}@fci.aun.edu.eg
[2] Institute of Computer Science, Friedrich-Schiller-University Jena, Germany
friederike.klan@uni-jena.de

ABSTRACT
As the number of bio-ontologies is increasing and growing very fast, we need a tool to support people in reusing them and selecting the most appropriate ones for their tasks. A number of ontology selection tools exist (e.g. BiOSS and NCBO's Recommender service), but they lack individual adaptations for preferences of their users. For that, we aim to develop a new ontology selection tool considering user preferences (i.e. the domain of interest, ontology reusing purpose, and preferred evaluation criteria), user interactions, feedback, and historical searches. In this paper, we describe our proposed framework aiming to add personalized support to the ontology selection process.

Categories and Subject Descriptors
H.1.m **[Information Systems]**: Models and Principles-*miscellaneous*; H.3.1 **[Information Storage and Retrieval]:** Content Analysis and Indexing—*thesauruses*; H.3.3 **[Information Storage and Retrieval]**: *information filtering, relevance feedback, and retrieval models, selection process.*

General Terms
Knowledge Representation, Knowledge Engineering, Knowledge Management.

Keywords
Personalized ontology selection, ontology evaluation, interactive feedback, user preferences.

1. INTRODUCTION
Ontologies are considered the core of the semantic web; they are mainly used in many different domains to represent knowledge, share, and reuse it. Using ontologies as a source of controlled vocabularies of information in different Bio-domains such as Biomedical, Bioinformatics, and Biodiversity, led to a large and steadily growing number of ontologies describing these domains. NCBO's BioPortal[1] and OBO foundry[2] are examples of ontology repositories that contain hundreds of Bio-ontologies available in several representation formats (i.e. OWL, RDF, OBO, etc.).

However, building an ontology from scratch is a complex process that consumes much time and effort and needs collaborative cooperation from both ontology engineers and domain experts. For that, ontology building is increasingly becoming a reuse-centric process, where existing ontologies are reused in order to develop new integrated ontologies for a certain application [1]. For better ontology reuse results, ontologies must be appropriately selected based on requirements of the targeted application area. Ontology search engines, such as Swoogle [2] and Watson [3], are used to search and retrieve ontologies according to user queries. But for ontology selection we need a step afterwards, we need to evaluate those retrieved (candidate) ontologies according to some evaluation criteria and select the most relevant set of ontologies [4].

Existing ontology selection tools (e.g. BiOSS [4] and NCBO's Recommender service [5]) allow users to input keywords or text, and search for candidate ontologies by matching the input keywords with ontology concepts. Next, candidate ontologies are evaluated using some fixed evaluation criteria, ranked, and outputted in the form of single ontologies, or combined set of ontologies. Although, they allow the user to control the importance of each evaluation criteria (by giving weights to them), they do not take other user preferences into account, such as, the domain of interest, the purpose of reusing ontologies, and the preferred evaluation criteria.

In this paper, we take the first step towards a personalized ontology selection tool that supports its users in reusing ontologies by recommending the most appropriate ontologies according to their preferences and requirements. The system models its users and builds user profiles by considering their explicit preferences along with their implicit feedback on the retrieved ontologies, and their historical searches. It iteratively improves the results until the user is satisfied.

The rest of the paper is organized as follows: in Section 2, we present the related work and use case study. Section 3 presents the framework of the proposed approach with detailed description of how it works. Finally, Section 4 concludes the paper and provides future research outlines.

29th GI-Workshop on Foundations of Databases (Grundlagen von Datenbanken), 30.05.2017 - 02.06.2017, Blankenburg/Harz, Germany.

[1] https://bioportal.bioontology.org/
[2] www.obofoundry.org/

2. USE CASE STUDY and SYSTEM REQUIREMENTS

In order to build a system that supports people in reusing ontologies, we first need to identify the requirements needed for such a system. For that we conducted an extensive survey for a number of case studies that reused ontologies in different domains and for different purposes. We compared the cases under study with respect to the following aspects: the domain of interest, the purpose of ontology reusing, and the methodology used to select and evaluate candidate ontologies. A detailed discussion is out of the scope of this paper and will be published elsewhere. In the following subsections we briefly summarize the main steps of the ontology reusing process and the results of the investigation, followed by a conclusion and a set of system proposed requirements.

2.1 Ontology Reuse Process

Ontologies are originally built to be shared and reused, and ontology reusing is an integral part of ontology engineering. When deciding to develop new ontology, especially for an interdisciplinary domain, it is a better choice to reuse existing ontologies rather than building it from scratch [10]. The ontology reuse process consists of 4 main phases, which are:

1. **Scope definition:** in this step the domain expert is responsible for defining the scope of domain under study and specifying the requirements. This step also includes searching for candidate ontologies.
2. **Ontology selection:** to select a set of ontologies to be reused by assessing and evaluating candidate ontologies across some evaluation criteria.
3. **Ontology integration and merging:** then, selected ontologies are translated into one representation format and integrated or merged together to generate new domain ontology.
4. **Ontology assessment and verification:** finally, the outputted ontology is assessed and verified.

2.2 Ontology Reuse: Domain and Purpose

Ontologies are reused in many different domains and for different purposes [10]. From our study of ontology reuse cases, we found that the bio-medical domain has the greatest number of cases [6]. Actually, this was expected as it has a large and growing number of ontologies with large group of audiences (i.e. UMLS[3], SNOMED-CT[4], etc.). Furthermore, we discover that people are reusing ontologies for different purposes, specifically for integrating existing ontologies to build new ontological knowledgebase [7, 8], and for helping to build ontology-based applications, such as natural language processing applications [9], information retrieval applications [10], or context-aware applications [11]. Moreover, we also found that a fully automatic ontology reuse process does not exist; it is either done manually or semi-automatically. In manual cases, ontology engineers along with domain experts are working together to manually select, integrate, and assess resulting ontology/application [7, 10]. In semi-automatic cases, the role of the domain experts differ from one case to another, but generally they perform one or two of the ontology reusing steps while the rest of steps are done using existing tools. For example, MetamorphoSys is used to select lung-pathology relevant ontologies from UMLS ontologies [10]. Other tools are used for searching for ontologies and translating them to different representation formats [8].

2.3 Evaluation Criteria

Ontology selection is the main step in the ontology reuse process. To select adequate ontologies one needs to search for candidate ontologies, which have high percentage of concepts that match input keywords (input coverage). Then, candidate ontologies are evaluated using some predefined evaluation criteria. According to M. Sabou [12], ontology evaluation is "the core" of the ontology selection process; it aims to assess ontologies across a set of predefined evaluation criteria in order to determine which would best meet the requirements. According to [14] ontologies are considered complex structures that need to be evaluated across different levels, such as lexical level, structure level, context level, or syntactic level. Coverage, size, popularity, semantic richness, representation format, formality, and uptodateness are examples of evaluation criteria used in ontology selection approaches. Many different categories of ontology-based evaluation criteria are surveyed in [15, 16].

From our study of ontology reuse cases, we found a number of commonly used evaluation criteria. For example, keyword coverage was the most widely used one, as it can filter out irrelevant ontologies based on the percentage of containment of the specified keywords. Also, popularity of ontologies is often used although it is not always clear how to assess it. For example, in BiOSS [4], popularity is measured using tags in social media portals. This could not be always an accurate method especially in scientific domains. Formality and consistency of ontologies, are also assessed to ensure reusing consistent ontologies with formal representations. Moreover, there are cases that needed to select ontologies in standardized representation language such as OWL [8], as they provide support for information integration, extensibility and flexibility to change [15].

2.4 Existing Ontology Selection Tools

As the main interest here is the ontology selection process, we need to get a closer look at the main functionalities of existing ontology selection tools to discover the missing requirement in such systems. Table 1 provides a brief comparison for existing ontology selection approaches. BiOSS [4] and NCBO's recommender system [5] are ontology selection web tools, they evaluate ontologies using different evaluation criteria. For example, BiOSS uses coverage, semantic richness, and popularity, while NCBO's recommender system uses coverage, acceptance, detail and specialization. JOYCE [16] is also a new web-demo aiming to select parts of ontologies (modules) instead of selecting the whole ontologies. This could ensure retrieving important parts of ontologies and avoiding irrelevant ones, but also needs to ensure all ontologies in the repository are modularized accurately. Coverage, overlap, and overhead are JOYCE's criteria to evaluate and select ontology modules. Groza [17], and Batet [18] evaluate and select ontologies based on analytic hierarchy process and semantic similarity, respectively. Most of these approaches are applied on the biomedical domain ontologies.

[3] https://www.nlm.nih.gov/research/umls/
[4] http://www.snomed.org/

Table 1. Comparison of existing ontology selection approaches

	Biomedical Ontology Selection System (BiOSS)[4]	NCBO's Recommender System 2.0[5]	Ontology Selection based on Analytic Hierarchy Process[17]	Ontology Selection for Semantic Similarity Assessment [18]	WebCore ontology Recommen -dation system[19]	JOYCE[16]
Domain	Biomedical	Biomedical	Tourism	Biomedical	General	Biomedical
Input	Keywords	Keywords or text	Keywords	Keywords	Keywords	Keywords or text
Output	Single ontology, or combinations	Single ontology, or combinations	Single ontology	Single ontology	Single ontology	Combined ontology modules
Ontology selection process characteristics:						
• **Iterative**	• No	• Yes	• No	• No	• No	• No
• **Interactive**	• No	• Yes	• No	• No	• No	• No
• **Personalized**	• No	• No	• No	• No	• Yes	• No
• **Adapted evaluation Criteria**	• No	• Yes	• Yes	• No	• Yes	• Yes
Evaluation Criteria	• Input coverage • Knowledge richness • Popularity	• Input coverage • Acceptance • Detail • Specialization	• Language expressivity • Domain • Coverage • Size • Consistency • Cohesion	Semantic similarity	• Correctness • Readability • Flexibility • Type of formality • Type of model	• Coverage • Overlap • Overhead

Input to such tools is a list of domain-related keywords, but NCBO's recommender system and JOYCE allow user to input text. BiOSS, NCBO's recommender system and JOYCE output single ranked ontologies or combined sets of ontologies. Combinations of ontologies can provide much more input coverage than single ontologies. WebCore [19] is considered the first attempt towards a personalized ontology recommendation. It presents a collaborative recommendation framework where user profiles are manually defined by users or automatically generated from other users' data. The main limitation here is that the approach is not interactive; they do not provide profile learning technique or user feedback to update user profile.

2.5 System Proposed Requirements

To conclude, from our study of ontology reuse cases and existing ontology selection tools, it becomes evident that tools for ontology selection do not provide a sufficient level of personalization, (i.e. are not flexible enough to be customized to their user's needs and requirements). They also don't allow for interactive selection processes based on feedback given by the user. For that we are proposing to develop an ontology selection tool providing personalized support to its users with the following requirements:

1. The tool is used for scientific purposes and includes ontologies from Bioportal and OBO foundry, which are in the biomedical, biological, bioinformatics, or biodiversity domains.
2. User can adapt evaluation criteria according to the purpose of reusing ontologies and adjust the importance (weight) of each criterion.
3. The tool should support domain experts in selecting and reusing ontologies and make their task easier, not replace them.
4. Allow different types of inputs such as keywords, textual resources, and possibly semantic relations.
5. It should be interactive (with its user) and iteratively improve the results according to user's preferences and feedback that are updated after each interaction.

3. PROPOSED FRAMEWORK

In this section we present our proposed personalized ontology selection framework, as shown in Figure 1. The framework consists of three main modules: the input manipulation module, the user modeling module, and the ontology selection module.

Before we go into the details of each module, we illustrate how ontologies, obtained from BioPortal and OBO foundry, are stored in the local Bio-ontology repository (requirement 1). Ontologies are downloaded and stored in a local repository, which in addition to the ontologies themselves also stores a "Meta-ontology". Meta-Ontology is an ontology providing vocabulary and semantics for the meta-information about the stored ontologies. Figure 2 presents a sample of the proposed meta-ontology as it describes the domain, available representation format, size, and type structure of the stored ontologies. This could help in classifying ontologies in the repository, and to speed up the recommendation process. For example, if a user prefers to select biomedical ontologies with small size and in OWL format, we just search for ontologies that match those criteria instead of searching the whole set of ontologies. The Meta-ontology was first introduced in [20],

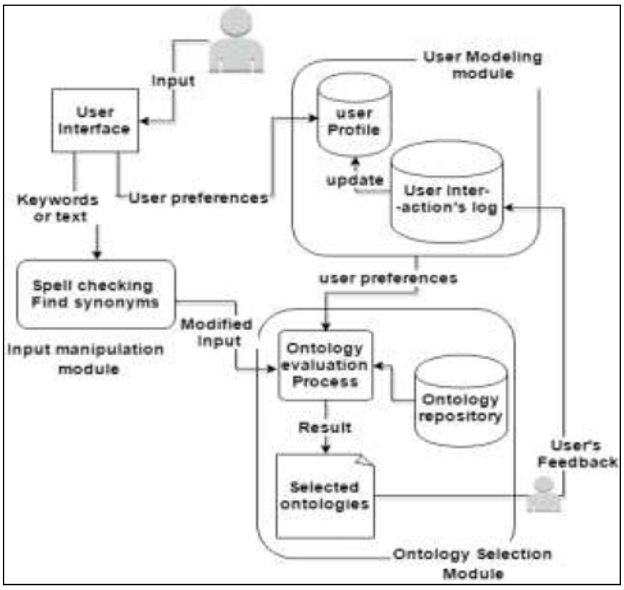

Figure 1. Proposed framework architecture

and previously used in different applications such as on quality assurance for ontologies [21] and for knowledge acquisition and exploitation[22].In the following subsections, we present details about the framework modules.

3.1 Input Manipulation Module

The first component of the architecture a user interacts with is the input manipulation module. It provides a simple form asking the user to input information about his/her domain of interest, purpose of reusing ontologies, preferred evaluation criteria, and input query (this could be a set of keywords or a piece of text). In the following subsections we illustrate how this works.

3.1.1. User Domain:

The user is allowed to select his preferred domain from a list of domains. For example, if the user selects the biomedical domain, the system should search for biomedical ontologies only.

3.1.2. Purpose of Reusing Ontologies:

Also, the system allows the user to select the intended purpose for reusing ontologies. From our survey, we found that main purposes for reusing ontologies are integration, ontological application development, knowledge representation, and others. According to the selected purpose, the system could suggest appropriate evaluation criteria. For example, if the user selects that he is reusing ontologies for ontology integration, then retrieved ontologies are preferably include most of the input keywords (high coverage percentage), are consistent, and possess formal representations in similar representation formats. Here, the system could suggest coverage, consistency, formality, and representation format, as the appropriate evaluation criteria. On the other hand, if the purpose of reusing was searching for ontologies, the system should not recommend evaluation criteria and leave it to the user to select the criteria that he is searching for (requirement 2).

Of course, adapting evaluation criteria according to reusing purposes is not an easy task, but it could be accomplished by conducting user's surveys, and analyzing historical searches.

3.1.3. Evaluation Criteria

We believe that ontologies are complex structures that need to be evaluated appropriately [23]. For that, the system provides criteria that evaluate them on the lexical, semantic, syntactic, structure, and context levels. Coverage, consistency, size, formality, popularity, uptodateness, representation languages, completeness, type structure are examples of criteria used for ontology evaluation. From the previous subsection, the system could give suggestions for the suitable criteria to the user that could be used, but the user has the right to select additional criteria or remove others. In the user interface form, evaluation criteria are listed in a check list, and the user can select any number of them.

3.1.4. Input Query

Finally, the user enters the input query in the form of keywords or a piece of text in a text field (requirement 4). User input is used for two different purposes; to explicitly provide information in order to generate user profiles (see Section 3.2) and to select appropriate ontologies.

So, in order to select the appropriate ontologies, the system allows the user to describe the desired domain either by entering a set of domain-related keywords or a piece of text, e.g. from publications. The keywords are manipulated using spellchecking tools, to check the spelling (to discover typing errors), and to find synonymous concepts (i.e., a synonym for the keyword Lithosphere is Geosphere).

The set of modified keywords will be the input to the ontology selection module (see Section 3.3).

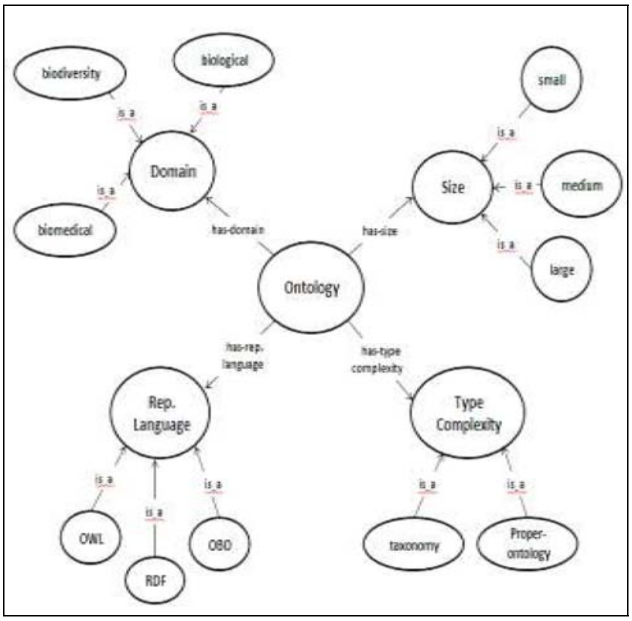

Figure 2. A sample from the meta-ontology

3.2 User Modeling Module

This module aims to enhance the selection process by adding a personalized support component, taking into account user preferences. And as user profile is a basic component in any personalized system, the user modeling module consists of two main tasks, which are:

1. **Create initial user profile:**
   The recommender system uses user input and preferences (domain of interest, purpose of reusing ontologies, and selected evaluation criteria) to initially create the user profile.
2. **Update user profile:**
   The user interaction log is a repository where user interactions and feedback are stored in order to update user's profile after each system interaction (requirements 5). Hybrid feedback is the source of information stored in the interaction log, which is obtained:
   - Explicitly from ratings provided by the user on the selected ontologies (if he/she is satisfied with the suggested ontologies or not).
   - Implicitly from historical searches and button clicks to navigate selected ontologies.

3.3 Ontology Selection Module

This module receives the modified keywords from the input manipulation module and user preferences from user recommendation module and begins the ontology selection process. At first, the process matches the keywords with concepts of the ontologies, taking into account explicit user preferences. This cuts the search space by excluding ontologies that do not meet user preferences. This outputs a set of candidate ontologies that are evaluated and ranked to be outputted to the user.

The output is displayed to the user who provides feedback explicitly by rating the output results, or implicitly by tracking his clicks on ontologies links which means that he needs to navigate or reuse this ontology. User feedback is stored in the interaction log repository and is used to update user profile. The process is repeated until the user is satisfied with the results.

4. CONCLUSION

The large number of existing ontologies in every field is motivating to reuse existing ones for different purposes such as building new ontologies, integrating them, or building ontological applications. But as ontologies now are very complex structures, it is not easy to select appropriate ontologies to be reused. A number of ontology selection tools exist and apply different evaluation criteria to assess and rank ontologies, but they lack to consider user preferences while selecting ontologies. In this paper, we present a proposed framework that aims to add personalized support to an ontology selection tool. Our future direction is to apply semantic technologies to create user profiles and give more consideration to user interactions and feedback that are used to update user profiles. Our focus is on bio-ontologies, using ontologies from NCBO's BioPortal and OBO Foundry.

5. ACKNOWLEDGMENT

This work is partly funded by the DAAD project "BioDialog". We are particularly grateful to Prof. Birgitta König-Ries, and Dr. Alsayed Algergawy.

6. REFERENCES

[1] Suárez-Figueroa, M. C., Gómez-Pérez, A., & Fernández-López, M. 2012. The NeOn methodology for ontology engineering. *Ontol. Eng. a networked world*, pp. 9–34.

[2] Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y., Sachs, J., Reddivari, P., Doshi, V. 2004. Swoogle : A Semantic Web Search and Metadata Engine. *In Proc. 13th ACM Conf. on Information and Knowledge Management*, pp. 10–1145.

[3] d'Aquin, M., Sabou, M., Dzbor, M., Baldassarre, C., Gridinoc, L., Angeletou, S., & Motta, E. 2007. WATSON : a gateway for the semantic web.

[4] Martínez-Romero, M., Vázquez-Naya, J. M., Pereira, J., & Pazos, A. 2014. BiOSS: A system for biomedical ontology selection. *Computer methods and programs in biomedicine*, 114(1), 125-140

[5] Martinez-Romero, M., Jonquet, C., O'Connor, M. J., Graybeal, J., Pazos, A., & Musen, M. A. 2016. NCBO Ontology Recommender 2.0: An Enhanced Approach for Biomedical Ontology Recommendation. arXiv preprint arXiv:1611.05973.

[6] Shah, T., Rabhi, F., Ray, P., & Taylor, K. 2014. A guiding framework for ontology reuse in the biomedical domain. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 2878-2887). IEEE.

[7] Coulet, A., Smaïl-Tabbone, M., Napoli, A., & Devignes, M. D. 2006. Suggested ontology for pharmacogenomics (SO-Pharm): modular construction and preliminary testing. In *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops* (pp. 648-657). Springer Berlin/Heidelberg.

[8] Caldarola, E. G., Picariello, A., & Rinaldi, A. M. 2015. An approach to ontology integration for ontology reuse in knowledge based digital ecosystems. In *Proceedings of the 7th International Conference on Management of computational and collective intElligence in Digital EcoSystems* (pp. 1-8). ACM.

[9] Peralta, D. N., Pinto, H. S., & Mamede, N. J. 2004. Reusing a time ontology. In *Enterprise Information Systems V* (pp. 241-248). Springer Netherlands..

[10] Simperl, E. 2009. Reusing ontologies on the Semantic Web: A feasibility study. *Data & Knowledge Engineering*, 68(10), 905-925.

[11] Poveda Villalon, M., Suárez-Figueroa, M. C., García-Castro, R., & Gómez-Pérez, A. 2010. A context ontology for mobile environments.

[12] Sabou, M., López, V., Motta, E., & Uren, V. 2006. Ontology Evaluation on the Real Semantic Web. *In Proceedings of the 4th Int. Workshop on Evaluation of Ontologies for the Web (EON'06), at the 15th Int. World Wide Web Conference (WWW'06)*. Edinburgh, UK.

[13] Hlomani, H., & Stacey, D. 2014. An extension to the data-driven ontology evaluation. In Information *Reuse and Integration (IRI), 2014 IEEE 15th International Conference on* (pp. 845-849). IEEE.

[14] Brank, J., Grobelnik, M., & Mladenic, D. 2005. A survey of ontology evaluation techniques. In *Proceedings of the conference on data mining and data warehouses (SiKDD 2005)* (pp. 166-170)..

[15] Reynolds, D., Thompson, C., Mukerji, J., & Coleman, D. 2005. An assessment of RDF/OWL modelling. *Digital Media Systems Laboratory*, HP Laboratories Bristol, 28.

[16] Faessler, E., Klan, F., Algergawy, A., König-Ries, B., & Hann, U. 2016. Selecting and Tailoring Ontologies with JOYCE. in *20th International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, 2016.

[17] Groza, A., Dragoste, I., Sincai, I., Jimborean, I., & Moraru, V. 2014. An Ontology Selection and Ranking System Based on the Analytic Hierarchy Process. In *Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2014 16th International Symposium on* (pp. 293-300). IEEE.

[18] Batet, M., & Sánchez, D. 2015. Ontology Selection for Semantic Similarity Assessment. In *ICAART (2)* (pp. 569-576).

[19] Cantador, I., Fernández, M., & Castells, P. 2007. Improving ontology recommendation and reuse in WebCORE by collaborative assessments.

[20] Cantador, I., Bellogín, A., & Castells, P. 2008. A multilayer ontology-based hybrid recommendation model. *Ai Communications, 21*(2-3), 203-210.

[21] Ochs, C., He, Z., Zheng, L., Geller, J., Perl, Y., Hripcsak, G., & Musen, M. A. 2016. Utilizing a structural meta-ontology for family-based quality assurance of the BioPortal ontologies. *Journal of biomedical informatics, 61*, 63-76.

[22] Lauras, M., Bénaben, F., Truptil, S., Lamothe, J., Macé-Ramète, G., & Montarnal, A. 2014. A meta-ontology for knowledge acquisition and exploitation of collaborative social systems. In *Behavior, Economic and Social Computing (BESC), 2014 International Conference on* (pp. 1-7). IEEE.

[23] Park, J., Cho, W., & Rho, S. 2010. Evaluating ontology extraction tools using a comprehensive evaluation framework. *Data & Knowledge Engineering, 69*(10), 1043-1061.