Local Ontologies Merging in Data Ponds

Jabrane Kachaoui
Information Technology and Modeling
Hassan II University
Casablanca, Morocco
jabrane2005@gmail.com

Abdessamad Belangour
Information Technology and Modeling
Hassan II University
Casablanca, Morocco
belangour@gmail.com

Abstract—**Today, Ontologies have a major place in knowledge representation and modeling. They are used to formalize a domain knowledge and add a semantic layer to current systems and applications. Ontologies make it possible to explicitly represent the knowledge of a domain by means a formal language so that they can be manipulated automatically and shared easily. They are widely used in various fields of research such as Knowledge Representation (KR) and Data Integration (DI). However, the effectiveness to interoperate learning objects among various learning object repositories is often decreased because of using different ontological schemes for annotating learning objects into every learning object repository. Hence, semantic heterogeneity and structural differences between ontologies need to be resolved so as to generate common ontology to expedite learning object reusability. This paper focused on automated ontology mapping and merging concept. The study significance lies in an algorithmic approach for mapping attributes of learning objects/concepts and merging them based on mapped attributes; identifying suitable threshold value for mapping and merging.**

Keywords—Ontology, Data Lake, Knowledge representation, Semantic, Data Ponds,

I. Introduction

The principle objective of current systems is to provide, from various data sources, information that is relevant to a user's request. Currently, a huge amount of information is available in all kinds of areas. This growing volume of information with an abundant production of data (scientific articles, information portals, reports, etc.) influences performance of these systems [1]. In addition, this information is generally expressed in Natural Language (NL) (in the form of texts) and therefore in an unstructured format, which makes their automatic processing difficult. Moreover, users can have levels of expertise and different profiles, ranging from specialists in treated field to general public users. In on the other hand, access to the right information is essential to help informed decision-making [2]. Faced with this challenge, it is necessary to design Retrieval Information (RI) tools more robust allowing users to easily find relevant information corresponding to their needs. These different questions have motivated lot of works in RI community, which led to the development of many models and tools, but also of advanced methodologies for their evaluation [3].

However, classic RI approaches are based on keywords; documents and queries are described by a set of words that they contain. Then, the correspondence between a document and a query is based on the number of terms they share. For that a document is selected, it must contain the same terms (or a part) as the user request. So the more the number of terms in common between the document and the request is higher, the more its relevance to this latter is higher. However, the relevant documents do not always contain the same terms as the request (problem of synonymy). Likewise, documents containing the same terms as the request is not necessarily relevant (homonymy and ambiguity issues). Thus, these approaches face two main linguistic problems: disparity and ambiguity of terms. Finally, classic RI generally considers terms indexing as independent entities and therefore does not considerate the relationships between these terms [4].

Semantic search, defined as search based on semantics of terms, was proposed to overcome these problems and improve the performance of RI traditional systems. The idea is to take into account the semantic content passed on by documents and requests rather than describing them by simple "word bags". Given the interest it has aroused, semantic RI has been the subject of numerous research works in recent years. Conceptual indexing used for explicitly describing documents by concepts has known a real passion. two categories has been distinguished: 1) statistical approaches which exploit the co-occurrence of terms in a corpus to define concepts using a reduction technique dimensions [5] and 2) resource-based semantics approaches, such as Thesauri [6] or Ontologies [7, 8], in which the concepts are explicitly defined. This work is part of the second approach and proposes to explore the potential of ontologies to guide an RI model in local Data Ponds. Indeed, ontology constitutes an ideal conceptual model to represent information but also to express information needs (requests). It also makes it possible to exploit the semantic relationships structuring these concepts to improve performance of RI systems. The reworking and expansion of initial user requests by semantically close concepts also allow improving responses relevance provided by the system. For example, for a user who wishes to access documents dealing with "Client", the system can propose documents also dealing with "Customer", by exploiting the specialization links [9]. In addition, ontology makes it possible to unify synonymous terms in different languages within the same notion (or concept). This is so particularly suitable and essential in a multilingual RI context.

The contributions of this paper are:
1. An algorithmic approach for mapping attributes from concepts and merging these concepts in local ontologies located in Data Ponds.

978-1-7281-8041-0/20/$31.00 ©2020 IEEE

2. Identification of an adapted threshold value for mapping and merging.

This paper is outlined as follows: First, related works in ontology mapping and merging are presented. This is followed by a discussion on problem statement. Next, an algorithmic approach for mapping and merging ontologies is described for merged ontology generation. Finally, concluding with conclusion and future works.

II. Related Works

Ontology engineering consists of the study of methods, techniques and tools to treat the different phases of development of an ontology. It is particularly interested in methodologies that provide guidance and techniques to support the process of ontologies construction. Given the importance of ontologies, several approaches have been proposed to guide this process, but there is none accepted by all [10].

Four main categories of approaches can be distinguished: construction ontologies approaches from scratch, construction ontologies approaches from texts, approaches based on the reuse of existing TORs (Terminological and Ontological Resource). Finally, approaches based on Crowdsourcing. Methodology varies depending on the use of ontology but also resources available to use for construction and merging. Application fields of ontologies being varied, approaches remain different and can be manual or semi-automatic. The first proposals in the field were manual methods where ontologies are built from scratch with help of domain experts. In the past of few years, ontological engineering has been marked through ontology merging approaches based on texts [11,12,13]. These approaches use tools as Natural Language Processing (NLP) methods. They are mainly intended to relieve the ontology construction process by automating certain steps using texts that are very important knowledge sources. Support development also made this textual data more accessible and thus facilitated their exploitation by NLP tools. Finally, facing challenges and taking into account required efforts for ontologies merging, the reuse of existing TORs is a very important question. Thus, new approaches have focused on reuse and reengineering of existing resources in ontological engineering [14,15,16].

This article focuses on presenting related works for ontology merging approaches based on texts and those based on Crowdsourcing.

A. Ontology merging appoaches based on texts

With advances in NLP, tools and methods for extracting knowledge from text have become robust. analysis and processing of textual documents becomes thus simple and easy. Thing that motivates development of ontologies merging methodologies from texts, is that they are considered as good sources of knowledge. These approaches often combine linguistic techniques, statistics and / or machine learning to extract ontological knowledge based on texts [17]. Linguistic techniques carry out a superficial or deep analysis of texts while statistical methods and machine learning use information such as frequencies documentaries of terms, their number of occurrences in the corpus and their co-occurrences. In practice, these two techniques are often combined. They are generally applied for domain ontologies merging. A preliminary step is therefore the constitution of a representative and adequate corpus of the field. Next, NLP techniques are used to process the corpus in order to extract some or all of the constituents of ontology: concepts, relationships, instances and axioms.

In [18],an iterative method for semi-automatic acquisition of ontology but also for the enrichment of existing ontologies is proposed. It provides a set of algorithms organized in modules allowing to extract primitives (concepts, relationships, etc.) **from texts. It's a method that combines NLP, machine learning** techniques and statistical methods. For extracting terms from ontology, a method based on statistical measures is applied to N-grams. A clustering method is then used to group these terms in concepts. Regarding the extraction of hierarchical relationships, relationships syntactic and expansion relations are used while the association rules technique makes it possible to acquire the non-hierarchical relationships. In this method, conceptualization is automatic; it allows generating an ontology automatically, this latter can then be refined and enriched with the help of an expert (adding new relevant concepts, deleting irrelevant concepts). This method was implemented in ontologies construction tool Text-To-Onto (Maedche et al., 2000).

Buitelaar and colleagues [19] have proposed a method mainly based on linguistics. It defines linguistic rules which allow extracting concepts and relationships from collections of linguistically annotated texts. It's an approach that integrates linguistic analysis into ontological engineering. it supports semi-automatic and interactive acquisition of ontologies from texts but also existing ontologies extension. This methodology is associated with an OntoLT plugin for Protégé [20] which remains the most used ontology editing tool Nowadays. OntoLT uses predefined matching rules that allow automatically extracting classes and indicate relationships from texts.

Another methodology was proposed and implemented in the OntoGen system [21,22] for semi-automatic ontologies merging. OntoGen aims to help user (often a domain expert) to identify primitives of ontology from a collection of documents. For this, it relies on techniques Latent Semantic Analysis (LSA) [23] and K-means clustering [24,25] to suggest concepts, relationships between these concepts and instances. It offers a graphical interface with several stands that allow user to visualize and explore concepts but also adjust ontology by adding new concepts or by editing existing ones.

B. Ontology merging appoaches based on Crowdsourcing

Unlike other approaches, new approaches increasingly using Crowdsourcing techniques [26, 27, 28] have appeared recently. Crowdsourcing consists of outsourcing tasks traditionally performed by a designated agent (such as an employee) by calling upon the intelligence and know-how of a large number of people [29]. Since the ontology merging process is tedious and generally requires a lot of time and resources, researchers have had the idea of involving a large group of users to simplify the task.

Mortensen et al [30] consider that Crowdsourcing can be a way to reduce difficulties raised by the development of large and complex ontologies. Thus, authors proposed methods based on crowdsourcing to achieve various ontological engineering tasks, such as assessing ontology quality and new ontologies production. In this approach, participants are subjected to a qualification test and those who take this test can access to tasks. The responses are then collected and evaluated. They show in their assessment that their method for the task of checking the ontology hierarchy gives an 82% accuracy.

Getman and Karasiuk [31] similarly proposed a method based on Crowdsourcing for ontology merging in the field of law. This method has been implemented and made available to a group of 20 users (students in law) for a semester. To simplify the task, each user worked on a subdomain. After the evaluation of two branches, (334 concepts structured by 338 relationships semantics) of the resulting ontology (6000 concepts) by experts in the field, they estimate the concept coverage for the branches analyzed at more than 90%. Although their results were conclusive, certain problems were raised after analysis. They noted by example that branches of the ontology created by different users are little connected. They also noticed that synonymous concepts were created separately. Authors conclude that involvement of qualified users in Crowdsourcing is beneficial for merging and maintaining ontology in a specific area such as law, but we must add a "polishing" task.

In addition to ontologies merging, crowdsourcing has also been used to enrichment [32] and alignment of ontologies [33]. For example, in [34], an ontology development approach based on Crowdsourcing is described. A system called OntoAssist was developed and used to integrate new concepts and semantic relationships between them in an ontology.

These different works have shown the potential of crowdsourcing techniques in ontological engineering. Using the knowledge of a large number of users qualified, thanks to these techniques, can therefore be complementary to conventional methodologies of merging by allowing lightening certain tasks. This distribution of tasks reveals the mutualisation notion and therefore of collaboration which is found in emerging approach [35].

This paper proposes an algorithmic approach blending the two discussed approaches above for ontologies merging. An Ontology merging approach based on texts is used based on data stored in Data Ponds and applying Crowdsourcing approach.

III. Backgound

The ontological semantic heterogeneity crop up from two scenarios. In the first one, ontological concepts are characterized with different terminologies (synonymy) in a given domain. For example, in Figure 1, terms client and customer, booking and reservation are synonymous but differently termed. In the second one, various signification are attributed to the same word in various contexts (polysemy). Furthermore, different taxonomies lead to structural heterogeneity between ontologies [36, 37, 38]. Examples of these differences in Figure 1 are between the concepts of airline and duration [39].

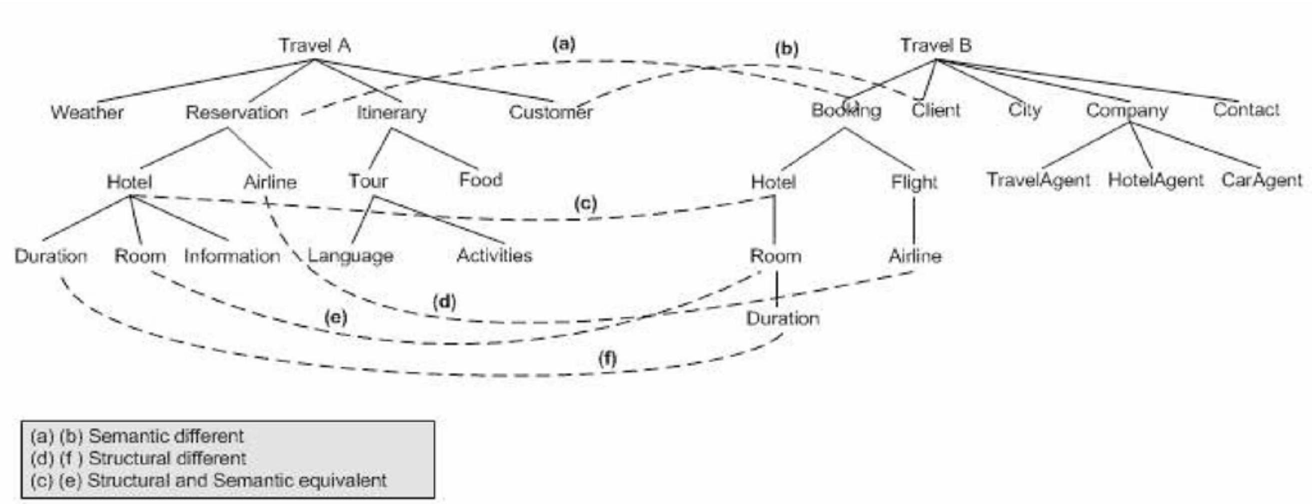

Fig. 1. Semantic and structural differences between ontologies

Therefore, there is a clear need for mapping and merging ontologies. Ontology mapping implies mapping structure and semantics for learning objects description in various repositories, while merging incorporates initial taxonomies through a shared schematic taxonomy.

Four major problems compel efforts around of ontology mapping and merging assignments [39]:

1. Semantic issues: When ontological concepts express a same domain with different terminology (polysemy and synonymy).thing that overlaps domains. To tackle with this issue, mapping tools are needed to perform metadata descriptions from semantic and lexical perspective to overcome polysemy and synonymy problems.
2. Structural issues: structure in mapping and merging ontology refers to structural taxonomy combining

concepts (objects). Various creators annotate learning objects with various ontological concepts. This makes syntactical issues, requiring merging tools that can capture several taxonomies and merge them into a shared taxonomy.

3. Mapping and merging scalability: This is particularly true for large ontological reference systems where Learning Objects repositories increase can be very big.
4. Prior knowledge lack: prior knowledge is necessary to map and merge ontologies using data mining methods. Thus, this knowledge is not always disposable. So unsupervised methods are necessary as an option in prior knowledge absence.

IV. GLOBAL ONTOLOGY CONSTRUCTION

A. Proposed approach

The evolution of data use contributes to the expansion of decision-making systems, integrating new data on a regular basis increases volume of data available in these systems. The concepts of active data or less active data appear. An organization data capital is therefore built up through the conservation of data in these systems. The notion of conservation and therefore archiving of this data also becomes an architectural element which must be taken into account in decision-making systems evolution.

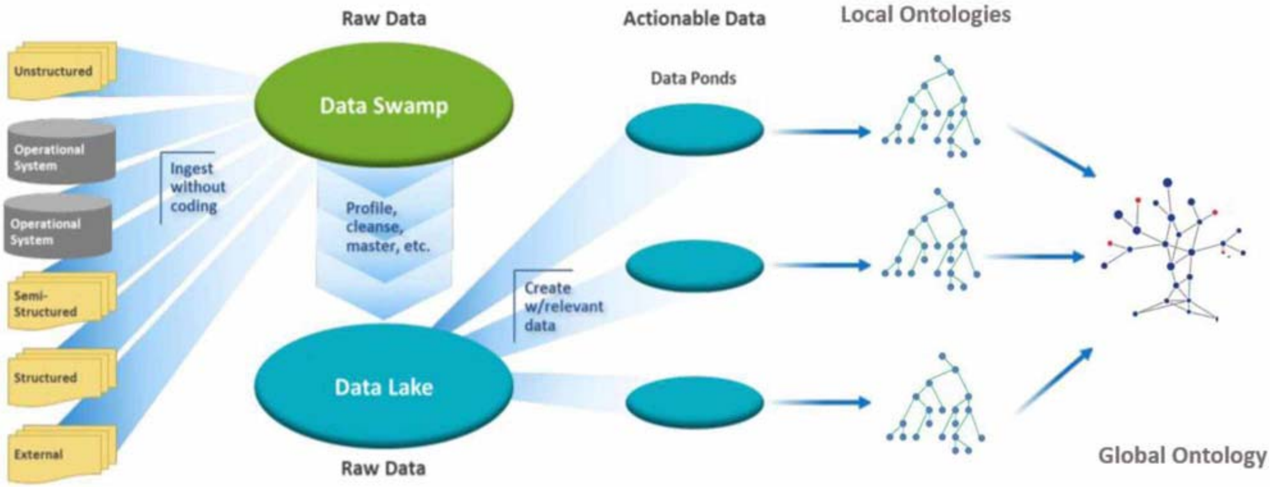

Fig. 2. Global system overview for data management through ontologies

When creating and representing relations and complex information between concepts, using semantic technologies can help to interpret information by identifying corresponding context. Semantic technologies can facilitate understanding data purpose and meaning (e.g. words, symbols, etc.) and complex concepts, as well as sharing knowledge between humans and machines.

It is necessary to develop a method of analyzing Big Data metadata which makes it possible to select data blocks from heterogeneous Big Data sources relevant to solve **user's** request. It should be keep in mind that annotations and task definition of Big Data are unstructured or semi-structured texts in NL. Consequently, their matching can be based on NL analysis methods but with Big Data ontology that contains knowledge on domain specificities and permits semantic processing of other elements of Big Data metadata. Creating a merging ontology based on local ontologies from data Ponds is also a part of this work.

What has been discussed before is illustrated in Figure 2 which represents the result of our previous works that carried out on building first pillars of this architecture [1], [2], [3], [25], [40], [41]. several stages have been approved concerning Data Lake and Data Warehouse complementarity, clustering stage using K-means based on metadata, transformation stage of unstructured data into structured data, and a last stage of global ontology merging that consists in facilitating data acquisition of pertinent information from various data ponds.

In this architecture, all data is stored in a Data Lake, this data is preprocessed, then analyzed to finally be routed to different Data Ponds following a well-defined algorithm combining K-means with metadata. Each Data Ponds has its own local ontology. Each ontology is a detailed description of data, which is used for the formal and declarative definition of its conceptualization.

B. Methodology

As already mentioned, ontologies merging is the mechanism of detecting the nearest semantic relationship between ontologies of relevant ontological entities such as attribute, concept, instance etc. The merged ontology is nevertheless the process of creating new ontologies of two or more current ontologies with similar or overlapping parts [42]. Given ontology A and ontology B as illustrated in figure 3, merged ontology is used to find out inherent semantics of ontological concepts between the ontologies A and B. Ontologies concepts may or may not be related. The process of merging ontologies examines semantics between ontologies and restructures concept-attribute relationships among and between concepts to merge ontologies.

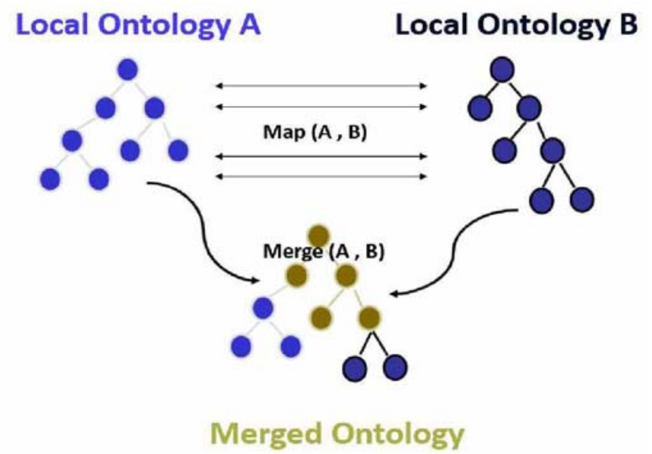

Fig. 3. Merged Ontology based on Local Ontology A and B

Ontology development is considered as specifying constraints, structure and data for further programs to use. System agents and other problem-solving concepts can utilize these ontologies as ready-to-use data that can be provided to the program for axioms and basic domain principles understanding. As discussed before in figure 3, for example ontology A and B, and a third ontology titled Merged Ontology comes from them. This new ontology makes interoperability easier between system agents based on ontologies A and B. The merged ontology can substitute old ontologies or used only as intermediary between systems using ontologies A and B. ontologies are mainly integrated in three manners depending required changes to derive the new ontology [43]:

1) By alignment

It is the lowest form for ontology integration. This needs small changes, but only handles limited interoperability types. It is used generally for RI, but does not support deep inferences. For instance, alignment maps relationships and concepts between ontologies A and B so that it partially preserves order of subtypes in the two ontologies. If an alignment maps a relation or concept x in ontology A with a relation or concept or y in ontology B, then x and y are equivalent. Concept mapping is not complete, so there may be a relation or concept in ontology A that has no equivalent in ontology B. Before both ontologies are aligned, it may be essential introducing new subtypes and supertypes of relations and concepts into one of the two ontologies. No other modification of axioms, calculations or definitions in A or B is made throughout alignment process [44].

2) By partial compatibility

Compared to alignment, this process is more interoperable, but also requires extensive changes. It is possible to be specified as an alignment of ontologies A and B that supports calculations or equivalent inferences on all equivalent relations and concepts. For instance, if two ontologies A and B are partially compatible, any calculation or inference that may be presented in one ontology using only aligned relations and concepts can be converted into equivalent calculation or inference in the other ontology.

3) By unification or total compatibility

This is also referred to as ontology merging. It gives full interoperability between ontologies data, but may need major changes. For instance, if partial compatibility of both ontologies A and B is expanded to entire compatibility in the new ontology, this new ontology involves all relations and concepts of the two ontologies A and B. Any calculation or inference that may be presented in one ontology may be mapped to an equivalent calculation or inference in the other ontology [45].

C. Proposal Algorithm

Terms and keywords used in the algorithm are first explained:

- Ontology O_A: Ontology O_A is the local data repository ontology A.
- Ontology O_B: Ontology O_B refers to local data repository ontology B.
- Formal context FC_A and FC_B: Formal context FC_A is the formal context representation of local ontology O_A conceptual relationship, meanwhile formal context FC_B is the formal context representation of local ontology O_B conceptual relationship.
- Reconciled formal context $_RFC_A$ and $_RFC_B$: Reconciled formal context $_RFC_A$ and $_RFC_B$ are formal context with normalized intents of local A and B ontological **concepts' properties.**

The prototypical implementation of the semi-automated mapping and merging process is explicated below:

Input : Two ontologies to be merged, O_A (local ontology A) and O_B (local ontology B)

Step 1: Ontological contextualization

The conceptual pattern of O_A and O_B is discovered using Formal Concept Analysis (FCA). Given an ontology **O:** = (OC, S_C, P, S_P, A), O_A and O_B are contextualized using FCA by respecting the formal context, FC_A and FC_B. The ontological concepts OC are denoted as J (objects) and the rest of ontology elements, S_C, P, S_P and A are denoted as M (attributes). The binary relation $\boldsymbol{R \subseteq J \times M}$ of the formal context denotes the ontology elements, S_C, P, S_P and A corresponding to the ontological concepts OC.

Step 2: Pre-linguistic processing

A similarity calculation, Levenshtein edit distance is applied to discover correlations between FC_A and FC_B attributes (ontological properties). The computed similarity value of ontological properties at or above threshold value is persisted in the context, else it appends into the context to reconcile the formal context, FC_A and FC_B. $_RFC_A$ and $_RFC_B$ are used as input for the next step.

Step 3: Contextual clustering

Self-Organizing Map (SOM) and K-means are applied to discover semantics of ontological concepts based on the conceptual pattern discovered in the formal context, FC_A and FC_B. This process consists of two phases:

(a) *Modeling and training*

Firstly, SOM is used to model the formal context $_RFC_A$ to discover the intrinsic relationship between ontological concepts of the source ontology O_A. Subsequently, k-means clustering is applied on the learnt SOM to reduce the problem size of the SOM cluster to the most optimal number of k clusters based on the Davies-Bouldin validity index.

(b) *Testing and predicting*

In this phase, new concepts from the local ontology O_B are discovered by SOM's Best-Matching Unit (BMU). SOM's BMU clusters formal concepts $_{R}FC_A$ into its appropriate cluster without need for prior knowledge of internal ontological concepts.

Step 4: Post-linguistic processing

The clusters, which contain ontological concepts of O_B, are evaluated by Levenshtein edit distance to discover semantic similarity between ontological concepts in the clusters. If the similarity value between the ontological concepts are at or above threshold value, the ontological concepts of O_B are dropped from the context (since they are similar to ontological concepts of O_A) and the binary relations $\boldsymbol{R \subseteq J \times M}$ are automatically updated in the formal context. Else, the ontological concept of O_B is merged with O_A. Finally, a Compounded Formal Context is generated.

Output : Merged ontology in a concept lattice is formed.

V. CONCLUSION

The emergence of the ontology merging technique is a means by which we overcome the restrictions and specificities of information and knowledge when it comes to a system that affects two (or more) similar domains. In this work, we have proposed an approach semi-automatic ontology merging. At this stage, human intervention is promising to validate results of similarity calculation module. This latter combines measure of lexical similarities, based on distance calculation between two strings describing both concepts in question, and semantics based on semantic enrichment of the two ontologies sources from an external "wordNet" resource. Next, the semantic similarity calculation between the two input concepts. Finally, each couple of concepts considered similar, by the combination of results obtained will be submitted to the Knowledge Engineer for validation, if accepted it will be replaced by a single concept result of elements merging of this couple, thus giving an ontology that covers a system domain.

This algorithm is far from complete, several improvements are to be completed to make it more efficient. In a future work, we aim to strengthen the performance of similarity identification module by using other RI techniques combined with **agents'** paradigm for building dynamic and automatic ontologies.

REFERENCES

[1] J. Kachaoui and A. Belangour. "Challenges and Benefits of Deploying Big Data Storage Solution". In Proc. New Challenges in Data Sciences: Acts of the Second Conference of the Moroccan Classification Societ, Article No.: 22, 2019, pp 1–5.

[2] J. Kachaoui and A. Belangour, "A Multi-criteria Group Decision Making Method for Big Data Storage Selection", In Proc. International Conference on Networked Systems, 2019, pp 381-386.

[3] J. Kachaoui and A. Belangour. "An Adaptive Control Approach for Performance of Big Data Storage Systems". In Proc. International Conference on Advanced Intelligent Systems for Sustainable Development, 2019, pp 89-97.

[4] A.Manepatil et al, "Keyword search in information retrieval and relational database system: Two class view". **Proc.** International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016.

[5] Scott.Deerwester et al. "Indexing by latent semantic analysis," Journal of the Society for Information Science, pp.41(6):391–407, 1990.

[6] Gonzalo, J., F. Verdejo, C. Peters et N. Calzolari. "Applying EuroWordNet to crosslanguage text retrieval". Computers and the Humanities 32 (2-3) : 185-207. 1998.

[7] A.Kiryakov et al, "Semantic annotation, indexing, and retrieval". Journal of Web Semantics, Volume 2, Issue 1, 2004, Pages 49-79.

[8] Castells et al., " An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval", IEEE Transactions on Knowledge and Data Engineering (Volume: 19 , Issue: 2 , Feb. 2007), 2007, 261 – 272.

[9] J .Kachaoui et al. Towards an Ontology Proposal Model in Data Lake for Real-time COVID-19 Prevention Cases. International Journal of Online and Biomedical Engineering (iJOE), 2020.

[10] **Lopez, F. "Overview Of Methodologies For Building Ontologies".** The Knowledge Engineering Review. Volume 17, Issue 2, June 2002 , pp. 129-156.

[11] Cimiano, P., Völker, J., 2005. Text2Onto A Framework for Ontology Learning and Datadriven Change Discovery. Proc. of International Conference on Application of Natural Language to Information Systems, NLDB 2005: Natural Language Processing and Information Systems pp 227-238.

[12] **Buitelaar, P., Olejnik, D., Sintek, M. "A Protege Plug**-in for Ontology **Extraction from Text Based on Linguistic Analysis". Proc. of the 1st** European Semantic Web Symposium (ESWS), 2004,pp 31-44.

[13] Nathalie.Aussenac-Gilles, Sylvie.Despres et Sylvie.Szulman, "The Terminae method and platform for ontology engineering from texts," Dans Proc. Conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge, pp. 199–223. IOS press. 2008.

[14] Simperl, E. "Reusing ontologies on the Semantic Web: A feasibility study". Data Knowl Eng 68, 905–925. doi:10.1016/j.datak.2009.02.002.2009.

[15] **Hepp, M., de Bruijn, J., "GenTax: A generic methodology for deriving** OWL and RDF-S ontologies from hierarchical classifications, thesauri, **and inconsistent taxonomies".** Semantic Web Res. Appl. 129–144, 2007.

[16] Villazón-Terrazas, B., Gómez-**Pérez, A., " Reusing and Re**-engineering Non-**ontological Resources for Building Ontologies", in: Suárez**-Figueroa, M. del C., Gómez-Pérez, A., Motta, E., Gangemi, A. (Eds.), Ontology Engineering in a Networked World. Springer, pp. 107–145, 2012.

[17] **Zouaq, A., Nkambou, R., "A Survey of Domain Ontology Engineering: Methods and Tools", in: Nkambou, R., Bourdeau, J., Mizoguchi, R.** (Eds.), Advances in Intelligent Tutoring Systems, Studies in Computational Intelligence. Springer Berlin Heidelberg, pp. 103–119, 2010.

[18] Maedche, A., Staab, S., "Ontology Learning for the Semantic Web". IEEE Intell. Syst. 16, 72–79. doi:10.1109/5254.920602,2001.

[19] Buitelaar, P., Cimiano, P., Magnini B. (Eds.), Ontology Learning from Text: Methods, Applications and Evaluation. IOS Press, pp. 3–12.2005.

[20] **Knublauch, H., Fergerson, R.W., Noy, N.F., Musen, M.A., "The Protégé** OWL Plugin: An Open Development Environment for Semantic Web **Applications", The Semantic Web –** ISWC 2004, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 229–243. 2004.

[21] Fortuna, B., Gro**belnik, M., Mladenic, D.,** "OntoGen: Semi-automatic Ontology Editor"**, Proc.** of the 2007 Conference on Human Interface: Part II. Springer-Verlag, Berlin, Heidelberg, pp. 309–318. 2007.

[22] Fortuna, B., M**ladenič, D.,** Grobelnik, M., "Semi-automatic Construction of Topic Ontologies", Semantics, Web and Mining, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 121–131. 2006.

[23] Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R., Indexing by latent semantic analysis. J. Am. Soc. Inf. Sci. 41, 391–407. 1990.

[24] **Jain, A.K., Murty, M.N., Flynn, P.J., "Data Clustering: A Review". ACM** Comput Surv 31, 264–323. doi:10.1145/331499.331504. 1999.

[25] J.Kachaoui and A.Belangour. Enhanced Data Lake Clustering Design based on K-means Algorithm. International Journal of Advanced

Computer Science and Applications, Volume 11 Issue 4, pp 547-554. 2020.

[26] Sarasua, C., Simperl, E., Noy, N.F., "CrowdMap: Crowdsourcing Ontology Alignment with Microtasks", The Semantic Web – ISWC 2012, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 525–541. 2012.

[27] Mortensen, J., Alexander, P.R., Musen, M.A., Noy, N.F., "Crowdsourcing Ontology Verification", Proc. of the 4th International Conference on Biomedical Ontology, ICBO 2013, Montreal, Canada, July 7-12, 2013, CEUR Workshop Proceedings. CEUR-WS.org, pp. 40–45. 2013.

[28] Getman, A.P., Karasiuk, V.V., "A crowdsourcing approach to building a legal ontology from text". Artif. Intell. Law 22, 313–335. doi:10.1007/s10506-014-9159-1. 2014.

[29] Howe, J., "Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business", 1st ed. Crown Publishing Group, New York, NY, USA. 2008.

[30] Mortensen, J., Alexander, P.R., Musen, M.A., Noy, N.F., 2013. "Crowdsourcing Ontology Verification", Proc. of the 4th International Conference on Biomedical Ontology, ICBO 2013, Montreal, Canada, July 7-12, 2013, CEUR Workshop Proceedings. CEUR-WS.org, pp. 40–45.

[31] Getman, A.P., Karasiuk, V.V., "crowdsourcing approach to building a legal ontology from text". Artif. Intell. Law 22, 313–335. doi:10.1007/s10506-014-9159-1. 2014.

[32] Lin, H., Davis, J., "Computational and Crowdsourcing Methods for Extracting Ontological Structure from Folksonomy", The Semantic Web: Research and Applications, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 472–477. 2010.

[33] Sarasua, C., Simperl, E., Noy, N.F., "CrowdMap: Crowdsourcing Ontology Alignment with Microtasks", The Semantic Web – ISWC 2012, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 525–541.2012.

[34] Lin, H., Davis, J., Zhou, Y., "Ontological Services Using Crowdsourcing". ACIS 2010 Proc. 2010.

[35] Tudorache, T., Nyulas, C., Noy, N.F., Musen, M.A., "WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the Web". Semant Web 4, 89–99. 2013.

[36] Euzenat, J. et al, "State of the art on current alignment techniques", http://knowledgeweb.semanticweb.org/. 2004.

[37] Noy, N. F., & Musen, M, "Algorithm and Tool for Automated Ontology Merging and Alignment". In Proc. of the 17th National Conference on Artificial Intelligence (AAAI'00), Austin, USA. 2000.

[38] Stumme, G., & Maedche, A. "FCA–Merge: A Bottom–Up Approach for Merging Ontologies". In IJCAI '01 - Proceedings of the 17th International Joint Conference on Artificial Intelligence, Morgan Kaufmann. USA. 2001.

[39] Ching-Chieh Kiu, Chien-Sing Lee, Ontology Mapping and Merging through OntoDNA for Learning Object Reusability, Educational Technology & Society, 2006.

[40] J. Kachaoui and A. Belangour. "MQL2SQL: A Proposal Data Transformation Algorithm from MongoDB to RDBMS", International Journal of Advanced Trends in Computer Science and Engineering, volume 9 No.2. pp 2457-2463. 2020.

[41] J. Kachaoui and A. Belangour. From Single Architectural Design to a Reference Conceptual Meta-Model: An Intelligent Data Lake for New Data Insights. International Journal of Emerging Trends in Engineering Research, volume 8 No.4. pp 1460-1465. 2020.

[42] Klein, M. "Combining and relating ontologies: an analysis of problems and solutions". In Proc of the 17th International Joint Conference on Artificial Intelligence (IJCAI-01), Workshop: Ontologies and Information Sharing, USA. 2001.

[43] J.F. Sowa et al. "Knowledge Representation: Logical". In: Philosophical, and Computational Foundations, Brooks/Cole, Pacic Grove, 2000.

[44] J. F. Sowa. "Building, sharing, and merging ontologies". In: web site: http://www.jfsowa.com/ontology/ontoshar.htm (2001).

[45] H. S. Pinto et al, "Some Issues on Ontology Integration". Proc. 16th International Joint Conference on Artificial Intelligence (IJCAI'99) Workshop: KRR5: Ontologies and Problem-Solving Methods: Lesson Learned and Future Trends", 2 August 1999, Stockholm, Sweden.