www.ijecs.in

International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issues 7 July 2016, Page No. 17317-17321

A Personalized Ontology Model For Web Mining Using Instance Matching

Madhuri S. Pawar[1], K. V. Metre[2]

[1]ME Computer Engineering, Student, MET Institute of Engineering, BKC, Nashik, Maharashtra, India
pawarmadhuri17@gmail.com

[2]Assistant Professor, MET Institute of Engineering, BKC, Nashik, Maharashtra, India
kvmetre@yahoo.co.in

Abstract: In rapid growth of internet, the amount of web information gathering becomes a challenging point for all users. Many existing retrieval systems have been developed to attempt to satisfy this problem. But still there is no complete solution to the problem. Ontology describes and formalizes standardized representation of knowledge as a set of concepts and the relationship between those concepts. In personalized web information gathering ontology is also used to represent the user profiles. For representing user profiles many existing models have been provided knowledge from either global or local knowledge base. The user background knowledge can be better discovered if we integrate global and local analysis. The proposed system emphasizes the specific semantic relation in one computational model. Ontology contains lots of instances. Automatically instance matching has become the fundamental issue. Instance matching approach is used to present based on discriminative property value. Ontology model in domain specific system with instance search gives more accurate result.

Keywords: Ontology, personalization, semantic relations, world knowledge, local instance repository, user profiles, web information gathering, instance matching

1. Introduction

Rapid growth and adoption of internet has further exacerbated user needs for information and knowledge location, selection and retrieval. Gathering the useful and meaningful information becomes challenging task to the users. The capture of user information needs is key point and user profile can help to capture information needs. User Profile reflects the interest of users. User profiles represent the concept models possessed by users and is implicitly generated from their knowledge description.

There are two way to represent user profile through global or local analysis. Global analysis uses existing knowledge representation like Word Net, digital libraries, and online categorization and Wikipedia. Global analysis techniques produce effective performance for user background description. Local analysis examine user local information or observes users behavior in user profiles. User background knowledge can be better discovered and represented if we can integrate global and local analysis within one model. In the proposed system the world knowledge description and a user's local instance repository (LIR) are used. World knowledge is commonsense knowledge acquired by people from experience and education. LIR is a user's personal collection of information items. This knowledge is used to gather relevant information about a user's preference and choices. A multidimensional ontology mining method include Specificity and Exhaustivity, is also introduced in the proposed model for analyzing concepts specified in ontologies.

In ontology instances includes a lot of valuable semantic information which used for matching task. Ontology contains large number of instances. Instance matching aims to link different instances that denote the same real-world object across heterogeneous data sources. For that purpose, used instance matching approach based on instance's discriminative property values. Firstly compare the discriminative property values for each instance. The instances which have similar discriminative value will be selected.

2. Related work:

2.1 Ontology learning:

Ontology learning is also known as ontology extraction which used to extract information. Ontology learning is used to extract a relevant concepts and relations from a given collection.Global knowledge were used by many existing information retrival system to learn ontologies for web information gathering. For example, Gauch et al. and Sieg et al. learned personalized ontologies from the Open Directory Project to specify users' interests in web search[3][4]. On the basis of the Dewey Decimal classification, King et al. developed IntelliOnto to improve performance in distributed web information retrieval[5]. Wikipedia was used by Downey et al. to understand underlying user interests in queries. These works effectively discovered user background knowledge description; however, their performance was limited by the quality of the global knowledge description. Many works mined user background knowledge from user local information[6]. Li and Zhong For ontology construction used pattern recognition and association rule mining techniques to discover knowledge[7].

2.2 Local profiles:

For capturing the user information needs user Profiles were used in web Information gathering. A user profiles reflect the interests of users. A profile can be used to store the description of the characteristics of person. User profiles are categorized into three groups: Interviewing, semi-interviewing, and non-interviewing. Interviewing user profiles are considered to be perfect user profiles. They are acquired by using manual techniques, such as questionnaires, interviewing users, and analyzing user classified training sets. One typical example is the TREC Filtering Track training sets, which were generated manually [4]. The users read each document and gave a positive or negative judgment to the document against a given topic.

Semi-interviewing user profiles are acquired by semi automated techniques with limited user involvement. These techniques usually provide users with a list of categories and ask users for interesting or non interesting categories. One typical example is the web training set acquisition model introduced by Tao et al. which extracts training sets from the web based on user fed back categories[5]. Non interviewing techniques do not involve users at all, but ascertain user interests instead. They acquire user profiles by observing user activity and behavior and discovering user background knowledge [6]. A typical model is OBIWAN, proposed by Gauch et al. which acquires user profiles based on users' online browsing history. The interviewing, semi-interviewing, and non interviewing user profiles can also be viewed as manual, semiautomatic, and automatic profiles, respectively[3].

2.3 Instance Matching:

There have been several approaches dealing with the instance matching problems. Instance matching approaches classified into two categories:

- Approaches based on instance properties classification:
  VMI in which instance properties classified in six categories: URI, Name, Meta, descriptive property values, discriminative property value and neighbours[13]. Z. Wang et al. classify the instance information in lexical information and structural information[16].
- Approaches based on interpretation of instance information:
  Existing works are based on the similarity strategies or techniques used to get more similar instances. For example, in COMA++, matching instances is based on two methods: content-based similarity which is based on string similarity functions like edit-distance and constraint-based similarity which is based on numerical or pattern constraints of the ontology [14]. In SIRIMI, matching process combines direct-base matching with a class-based matching technique [15].

3. Personalized Ontology Construction:

Personalized ontologies that formally describe and specifies user background knowledge. For example if we are searching for the subject "Europe", business travelers may expect different search from leisure travelers. A user may become a business traveler when planning for a business trip, or a leisure traveler when planning for a family holiday. A user's concept model may change according to different information needs.

3.1 Global Knowledge Representation

World knowledge is important for information gathering. World knowledge is common sense knowledge possessed by people and acquired through their experience and education. User background description is extracted from relevant and non-relevant concepts. User background knowledge is extracted from a world knowledge base encoded from the dynamic data set. We first need to construct the world knowledge base. The world knowledge base must cover an exhaustive range of topics, since users may come from different backgrounds. The dynamic data set was developed for organizing and retrieving information from a large volume of collections. The dynamic data set represents the natural growth and distribution of human intellectual work, and covers comprehensive and exhaustive topics of world knowledge.

3.2 Construction for Ontology Learning

The personalized ontologies represent the concept models possessed by users. Ontologies also dealing with a given topic. The subjects of user interest are extracted form WKB through user interaction. Ontology Learning environment(OLE) is developed for user interaction. For a given topic there are three sets of concepts for interesting subjects:

- positive subjects refer to the concepts that are interesting to the user or concepts relevant to the information need with respect to the topic.
- negative subjects refer to the concepts that may make paradoxical or ambiguous interpretations of the topic.
- neutral subjects refer to that have no indication of either positive or negative subject.

OLE provides users with a set of candidates to selects positive and negative subjects for given topic. These subjects are extracted from the WKB.

3.3 Instance Matching process

Approach is based on the instance properties classification. Instance information are distinguished into two types: discriminative property values and descriptive property values.

- The discriminative property values of instances are the characteristics of the instances which can be used directly to distinguish them.
- The descriptive property values are the descriptions of an instance.

In this process the candidate selection is based on the discriminative property values and the refinement result based on descriptive property values. When the result is obtained, two types of link are established: SameAs and ViewSameAs. process starts with taking two ontologies : source ontology Os and target ontology Ot as input. It consist of five main stages as shown in figure 1:

a. Preprocessing

All properties of the both ontologies will be extracted. It also collect information of each instance.

b. Property Classification

There are two types of properties are distinguished: discriminative properties and descriptive properties. The discriminative property values of instances are the characteristics of the instances which can be used directly to distinguish them. Some properties can be selected automatically.

c. Primary candidate selection based on discriminative properties

Primary candidate selection based on discriminative property value. For each instance select discriminative property pair. Then find similar pairs and result of this stage is used in the next one.

d. Result refinement using descriptive properties

Instance obtained in previous stage which has similar discriminative property values will be selected to compare their descriptive property values. Then compute the similarity of each descriptive property pair.

e. Combined Result

This stage combined the result of two previous stages. i.e. discriminative and descriptive property pairs.

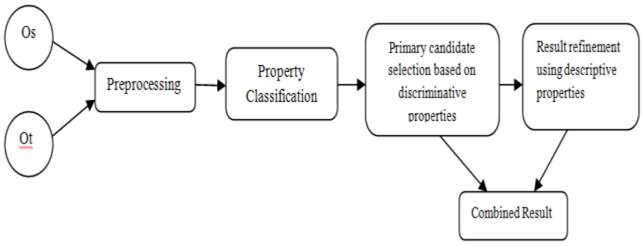

Os
Ot
Preprocessing
Property Classification
Primary candidate selection based on discriminative properties
Result refinement using descriptive properties
Combined Result

Figure 1: Instance Matching Process

3.4 PROPOSED MODEL:

The proposed ontology model aims to discover user background knowledge and learns personalized ontologies to represent user profiles. Figure 2 shows the proposed ontology model. A personalized ontology is constructed, according to a given topic. In the proposed Ontology model there are two types of search operations are performed i.e. local and global search. For global search it considers about the world knowledge base description. For local search it considers only about the local information. For better discovered and representation integrate both global and local search within one model. The world knowledge base provides the taxonomic structure for the personalized ontology. The user background knowledge is discovered from the user local instance repository. It retrieves global information based on the local database because of this time consumption for execution is very less and it gives accurate results, cost is also reduced. It covers wide range of topics.

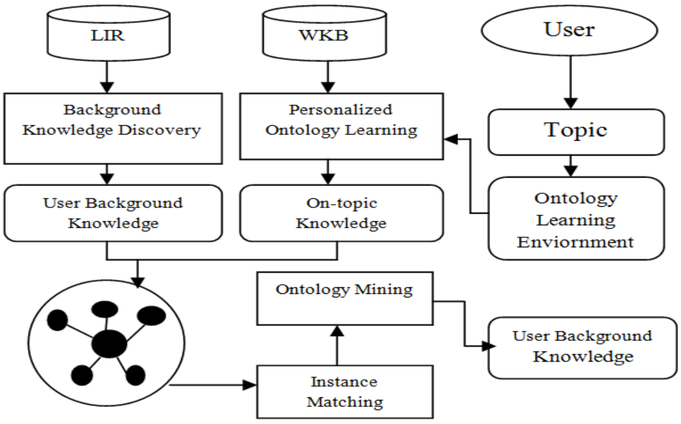

LIR
WKB
User
Background Knowledge Discovery
Personalized Ontology Learning
Topic
User Background Knowledge
On-topic Knowledge
Ontology Learning Enviornment
Ontology Mining
User Background Knowledge
Instance Matching

Figure 2: Proposed Model

3.5 Algorithm: Analyzing the Semantic Relation:

This algorithm is used to find semantic relation.

Input :

a personalized ontology 0(T) := (tax^s, vet); a coefficient θ between (0,1).

Output:

```
spe_a(s) applied to specificity.
1) set k = 1, get the set of leaves So from tax$^S$, for ($s_o \in S_o$) assign $spe_a(s_o) = k$;
2) get S' which is the set of leaves in case we remove the nodes $S_o$ and the related edges from tax$^s$;
3) if (S' == $\theta$) then return
4)  for each s' $\in$ S' do
5)  if (isA(s') == $\theta$) then $spe^1{}_a(s') = k$;
6) else $spe^1{}_a(s') = \theta$ x min{$spe_a$(s|s $\in$ isA(s'}};
7) if (part0f(s') == $\theta$) then $spe^2{}_a(s') = k$;
8) else $spe^2{}_a(s') = \frac{\sum s \in partof(s')spe_a(s)}{|partof(s')|}$
9) $spe_a(s') = \min(spe_a{}^1(s'), spe_a{}^2(s'))$;
10) end
11)  k = k x $\theta$, So = So U S' , go to step 2.
```

4. Methodology:

Ontologies are formal description for knowledge and specification of conceptualization. Ontologies are an important role in the semantic web and web information gathering. A world knowledge base is a global ontology that formally describes and specifies world knowledge. Here a user's background knowledge is extracted, including relevant and non-relevant to user information needs. The primitive concept is constructed based on the subjects. Subjects in the world knowledge base are linked to each other by semantic relations of is-a, part-of and related-to. An ontology is then constructed for the given topic using these user feedback subjects. The semantic relations linking the subjects in the WKB. It contains three types of knowledge like positive subjects, negative subjects and neutral subjects for a given topic. The personalized ontologies are formally defined :

1. Let S be a set of subjects, an element s $\in$ S is formalized as a 4-tuple s : < label, neighbor, ancestor descendant>
2. Let IR be a set of relations, an element r $\in$ IR is a 2-tuple r :<edge, type>
3. Let WKB be a world knowledge base, which is a taxonomy constructed as a directed acyclic graph. The WKB consists of a set of subjects linked by their semantic relations, and can be formally defined as a 2-tuple WKB :<S, IR>
4. The structure of an ontology that describes and specifies topic T is a graph consisting of a set of subject nodes. The structure can be formalized as a 3-tuple O(T):<S, tax^{S,} rel>

Ontology mining discovering the concepts , semantic relations, and instances in ontologies. In the multidimensional ontology mining methods: Specificity and Exhaustivity. Specificity describes the focus of subject's on a given topic. Exhaustivity restricts a semantic space dealing with the topic. This method aims to investigate the subjects and the strength of their associations in an ontology. We argue that a subject's specificity has two focuses: 1) on the referring-to concepts and 2) on the given topic. These need to be addressed separately.

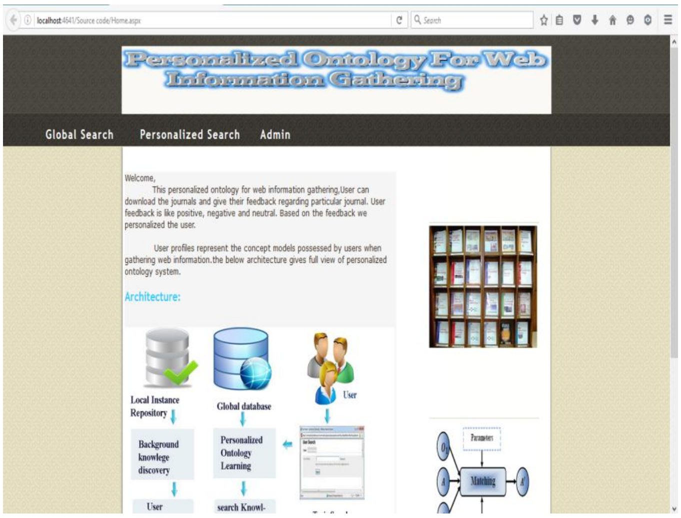
DOI: 10.18535/ijecs/v5i7.24

5. Results:

The experiments were conducted to compare the results generated by ontology model. In proposed model the world knowledge and a users local instance repository (LIR) are used. In ontology model the local profile is used to search which can bring out precise information based on the user's profile. Ontology mining discovers the concepts, semantic relations, and instances in ontologies.

5.1 Experiment :

The Ontology model has been implemented based on local, global database and based in figure 3 on the semantic relations in .NET framework. To evaluate our personalized ontology model the experiment are conducted on two dynamic data sets e.g. Educational ontology.The results are,

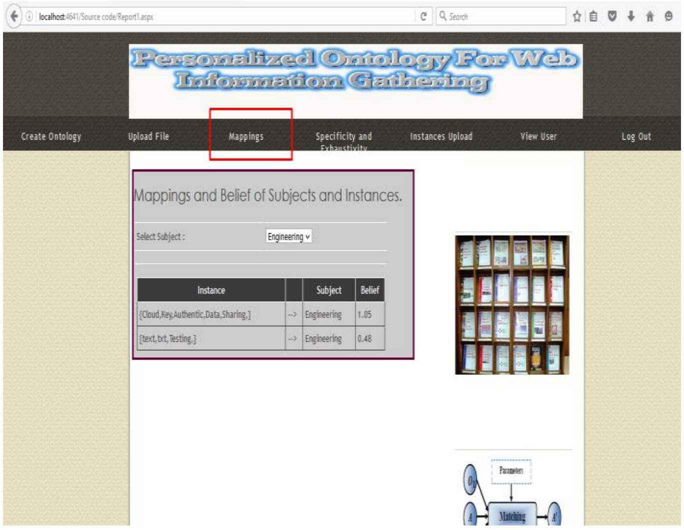

localhost:4641/Source code/Home.aspx
Search
Personalized Ontology For Web Information Gathering
Global Search
Personalized Search
Admin
Architecture:
Local Instance Repository
Global database
User
Background knowlege discovery
Personalized Ontology Learning
User
search Knowl-
Parameters
Matching
A
A'

Welcome,
This personalized ontology for web information gathering,User can download the journals and give their feedback regarding particular journal. User feedback is like positive, negative and neutral. Based on the feedback we personalized the user.

User profiles represent the concept models possessed by users when gathering web information.the below architecture gives full view of personalized ontology system.

Figure 3: Proposed Ontology Model

In this figure 4 describes about the reference strength between an instance and a subject. Each subject deals with only a part of the instance.


localhost:4641/Source code/Report1.aspx
Search
Personalized Ontology For Web Information Gathering
Create Ontology
Upload File
Mappings
Specificity and Exhaustivity
Instances Upload
View User
Log Out
Parameters
Matching
A
A'

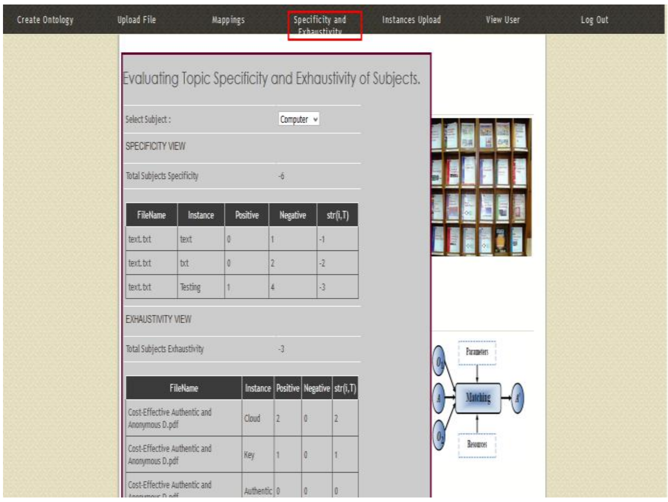
Mappings and Belief of Subjects and Instances.

Select Subject : Engineering

Instance		Subject	Belief
{Cloud,Key,Authentic,Data,Sharing,}	-->	Engineering	1.05
{text,txt,Testing,}	-->	Engineering	0.48

Figure 4 Mappings of Subjects and Instances

In figure 5 subjects are considered interesting to the user only if their specificity and exhaustivity are positive. A few theorems are introduce to restrict the utilization in ontology mining. Here theorem describes the leaf subject in terms of specificity and exhaustivity. If two subjects hold the same strengths to topic, then at a lower level must be more specific than the other one. It constrains the influence of positive and negative subjects to exhaustivity. Based on these, the definitions of specificity and exhaustivity are suitable for ontology mining.



Evaluating Topic Specificity and Exhaustivity of Subjects.

Select Subject : Computer

SPECIFICITY VIEW

Total Subjects Specificity -6

FileName	Instance	Positive	Negative	str(i,T)
text.txt	text	0	1	-1
text.txt	txt	0	2	-2
text.txt	Testing	1	4	-3

EXHAUSTIVITY VIEW

Total Subjects Exhaustivity -3

FileName	Instance	Positive	Negative	str(i,T)
Cost-Effective Authentic and Anonymous D.pdf	Cloud	2	0	2
Cost-Effective Authentic and Anonymous D.pdf	Key	1	0	1
Cost-Effective Authentic and Anonymous D.pdf	Authentic	0	0	0

Figure 5: Evaluating Topic Specificity and Exhaustivity of Subjects

5.2 Performance Analysis:

Features	Topic coverage	Accuracy	Time	Precision	Recall
TREC	0.5	1	0	1	0
Web	1	0	0.5	0	1
Category	0.51	1	1	1	0
Ontology	1.5	1.5	0.75	1	0.5

Table 1: Comparison Between Ontology And Existing Systems

The result are demonstrated in figure 6. As expected the performance of ontology model using instance matching is better than existing systems.

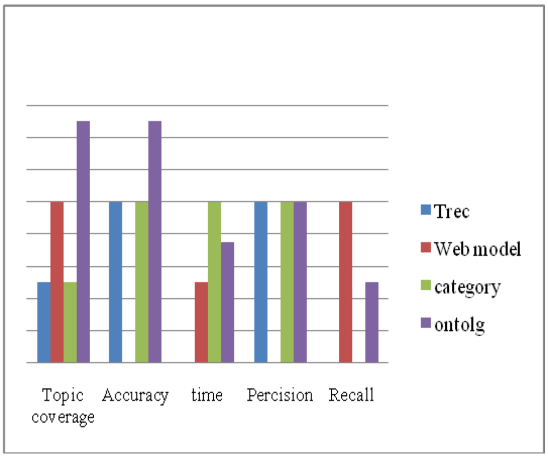


Figure 6: Comparative Results

Conclusion:

The proposed ontology model supplies a answer to emphasizing global and localized information in a single computational form. The outcome in this paper can be directed to the conceive of world wide web data gathering systems. The model furthermore has comprehensive contributions to the fields of Information Retrieval, World Wide Web understanding, Recommendation Systems, and Information schemes. In our future work, we will enquire the procedures that generate user localized example repositories to agree the representation of a international knowledge groundwork. The present work supposes that all client local instance repositories have content-based descriptors referring to the topics, although, a large volume of documents existing on the world wide world wide world wide web may not have such content-based descriptors. For this difficulty, the strategies like ontology mapping and text classification or clustering were proposed. These schemes will be enquired in future work to explain this difficulty. The investigation will continue the applicability of the ontology form to the majority of the existing World Wide Web articles and boost the assistance and significance of the present work.

References

[1] X. Tao, Y. Li, and N. Zhong, “A Personalized Ontology Model for Web Information Gathering", IEEE Transactions on Knowledge and Data Engineering , 2011.

[2] Wafa Ghemmaz, Fouzia Benchika, “Instance Matching Based on Discriminative Property Value’’, 5th International Conference on Information and Communication Technology and Accessibility (ICTA), Dec. 2015.

[3] X. Tao, Y. Li, and N, Zhong, “A Knowledge-based Model Using Ontologies for Personalized Web Information Gathering" Web Intelligence and Agent Systems, an International Journal, 2010.

[4] S. Gauch, J. Cha_ee, and A. Pretschner, “Ontology-Based Personalized Search and Browsing", Web Intelligence and Agent Systems , vol. 1, nos. 3/4,, 2003.

[5] A. Sieg, B. Mobasher, and R. Burke, “Web Search Personalization with Ontological User Pro_les", Proc. 16th ACM Conf. Information and Knowledge Management (CIKM 07), 2007.

[6] J.D. King, Y. Li, X. Tao, and R. Nayak, “Mining World Knowledge for Analysis of Search Engine Content", Web Intelligence and Agent Systems, vol. 5, no. 3, 2007.

[7] D. Downey, S. Dumais, D. Liebling, and E. Horvitz, “Understanding the Relationship between Searchers Queries and Information Goals", Proc. 17th ACM Conf. Information and Knowledge Management (CIKM 08) , 2008.

[8] Y. Li and N. Zhong, “Mining Ontology for Automatically Acquiring Web User Information Needs’’, IEEE Trans. Knowledge and Data Eng,vol. 18, no. 4 Apr. 2006.

[9] K. van der Sluijs and G.J. Huben, “Towards a Generic User Model Component’’, Proc. Workshop Personalization on the Semantic Web (PerSWeb 05), 10th Intl Conf. User Modeling (UM 05), 2005.

[10] R. Gligorov, W. ten Kate, Z. Aleksovski, and F. van Harmelen, “Using Google Distance to Weight Approximate Ontology Matches’’, Proc. 16th Intl Conf. World Wide Web (WWW 07), pp. 767-776, 2007.

[11] X. Tao, Y. Li, N. Zhong, and R. Nayak, “Automatic Acquiring Training Sets for Web Information Gathering’’, Proc. IEEE/WIC/ ACM Intl Conf. Web Intelligence, pp. 532-535, 2006.

[12] E. Frank and G.W. Paynter, “Predicting Library of Congress Classifications from Library of Congress Subject Headings’’, J. Am. Soc. Information Science and Technology, vol. 55, no. 3, pp. 214-227, 2004.

[13] R. Navigli, P. Velardi, and A. Gangemi, “Ontology Learning and Its Application to Automated Terminology Translation’’, IEEE Intelligent Systems, vol. 18, no.1, pp. 22-31, Jan. 2003.

[14] J. Li, Z Wang, X Zhang and J. Tang "Large scale instance matching via multiple indexes and candidate selection". Knowledge Based Systems, 2013. 50:p. 112-120.

[15] D. Engmann, S. MaBmann,Instance matching with COMA++, in proceedings of Datenbanksysteme in Buisiness,Technologie and Web(BTW 07), 2007, pp. 28-37.

[16] Samur Araujo, Due Thanh Tran, Arjen P.de Vries, and Daniel Schwabe, SERIMI: Class-Based Matching for instance Matching Across Heterogeneous Database. IEEE TRANSACTIONS ONKNOWLEDGE AND DATA ENGINEERING.VOL. 27.NO.5, MAY 2015.

[17] Z. Wang, J. Li, Y. Zhao, R. Setchi and J. Tang,A unified approch to matching semantic data on the Web Knowledge-Based Systems,2013. 39:173-184.