Asian Journal of Information Technology 15 (4): 840-845, 2016
ISSN: 1682-3915
© Medwell Journals, 2016

An Attempt for Content Based Matching on Semantic Web Using Relation Map Based Algorithmic Approaches

[1]S. Raja Ranganathan, [2]M. Marikkannan and [2]S. Karthik
[1]Department of Computer Science and Engineering, SNS College of Technology,
Coimbatore, Tamil Nadu, India
[2]Department of Computer Science and Engineering,
Institute of Road and Transport Technology, Erode, Tamil Nadu, India

Abstract: In semantic web, the information flow obtained from different relations is certain and processing those data across the relations are not easy without proper understanding about the semantic mapping between them. It is a complex process to manually identify these mappings and it is not possible over the web. It is required to develop tools for supporting relation mapping for the success of the semantic web. A technique named sealant is designed for machine based learning for identifying the mappings. For two given catalogs, the percept in one relation is identified by sealant and it predicts the most common percepts in other catalogs. A probability based explanations for many resemblance measures are viewed using sealant which works well with all of them. Furthermore, the sealant employs different learning techniques each of which utilizes several information types either in the data occurrence or in the catalog framework of the relations. The matching precision can be enhanced by expanding the sealant for integrating sound understanding and domain restrictions into the matching process. The technique varies with its working ways using clearly explained resemblance perception and effective integration of several types of understanding. The sealant is expanded for identifying difficult mappings between the relations and explains the analysis for its effective utilization.

Key words: Semantic web, relations, sealant, probability and catalogs

INTRODUCTION

The world wide web is experiencing >1.5 billions of web pages every now and then due to increasing requirements of the mankind. It creates a big issue for the software agents to analyze, recognize and understand the procedures for processing the information resulting in most of the web remains unused. As a solution the scholars have designed semantics which offers a structure and relations for describing the data semantics. These relations allow the software agents to better understand the relations so that it can potentially identify and aggregate the data for serving different tasks. For illustration it is required to obtain information about a person took a trip on a bus ‘A’. The only prevailing information about the person is his name ‘XYZ’ and he works for an organization ‘DEF’ but the branch he is working for is unknown. But, it was obtained that the person ‘XYZ’ has shifted himself from ‘DEF’ to ‘GHI’ where he is working for ‘MNO’. By utilizing the present world wide web it is not eazy to identify the person ‘XYZ’. All the information is restricted to a specific web page so a keyword based search would not be suitable. For addressing the problem semantic web would help greatly where a distinct indexing service makes the software agents to locate the closest available branch of the organization ‘DEF’ because the organization has some specific information obtained using a few associations as depicted in Fig. 1.

All the information is ordered into a catalog including team, designation and project. The project contains elements like team leader, co-workers and role in that project. This information allows the software agents to find that person with name ‘XYZ’ and analyzing the element ‘current project’ the software agent quickly locates that particular person within that region. For aggregating information from different relations, it is necessary to be aware about the semantic connections between their elements.

Literature reviw: Mohammad Mustafa addressed the semantic web as a development in current day web which indicates that the information is more extensively used by humans, computers, etc. (Agresti, 1990). It allows the

Corresponding Author: S. Raja Ranganathan, Department of Computer Science and Engineering, SNS College of Technology, Coimbatore, Tamil Nadu, India

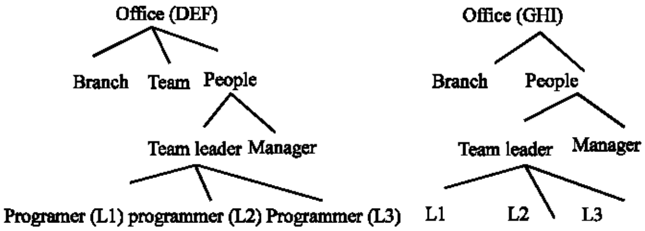

Fig. 1: Relations for XYZ

representation of the content and services in machine understandable form and allows clarification, investigation, sharing, advertising and collection for computerization. The design in based on the associations to reinforce the semantic web. The present day web is translated from machine decipherable to machine understandable form.

Ian Horrocks described that semantic web permits wide range of web reachable information and service for the access by both the human and computerized tools (Brickley and Guha, 2000). For aiding this process, RDF and OWL were designed for distributing and aggregating information and understanding called the relations. The languages and tools designed for sustaining them have quickly became the necessary standards for developing the relations and exploitation.

Li Ding explained that the semantic web is a standard and effective structure for improving the perception of understanding on web (Broekstra *et al*., 2002). The fundamental for semantic web is relations which clearly signify the concepts. The relation within semantic web is supported by languages like RDF, RDFS and OWL (Berners-Lee *et al*., 2001). The purpose is to analyze the prerequisite of the relations in terms of web and analyzes the above three languages with the conventional information demonstration and assessment tools for organizing the relations (Berlin and Metro *et al*., 2002).

Proposed solution: The disputes in identical associations on the semantic web could be addressed by devising a sealant technique which shares the machine based learning schemes to a moderately computerized semantic associations. The directory is the basic components for associations where the focus is to locate single association between the catalogs of the given catalogs for each perception node in a catalog and identify the most common perception node in the other catalog.

Connection definition: Initially, the connection between the two perceptions is addressed. It is noted that different meaning are available for connection each fits to certain situations. The aim is to study several practical measures for connection based on the probability joint distribution of the perceptions. Rather assigning to a exact meaning of connection the sealant estimates the probability joint distribution of the perceptions and allows the web applications to utilize the distribution for calculating any related connections. For instance let X and Y be the probability joint distribution containing of where:

$$P(X,Y), P(X,\bar{Y}), P(\bar{X},Y), P(\bar{X},\bar{Y})$$

the term $P(X,\bar{Y})$ is the possibility that a request in a domain belongs to perception X but not to Y. Using this an application can possibly describe the connection to be a fitting function of these four values.

Calculating connections: The further dispute for addressing is estimating the probability joint distribution for any two given discernment X and Y. Based on the distinct universal theory a terms such as P of X and Y can be fairly precise as the fraction of data request belonging to both. The issue reduces the assessment of each information percepts if it belongs to X∩Y It is viewed that the input for the problem encloses he discernment of X and discernment of Y in partition. The sealant attempts this issue using machine based learning methods by utilizing the perception of X to study a separation for X, then it catalogs percepts of Y to that separator and vice-versa for which X∩Y is available for recognizing the percepts.

Multiple approach learning: The usage of machine based learning into the perception might arise a question about which algorithm to be utilized for development. There are several types of information supplies for the perception categorization like its identifier, formats for value for its best utilization by the several different learning algorithms. The sealant uses a multiple approach learning to a set of learners for aggregating their calculation using an expert.

Developing domain restrictions: The sealant method also efforts to utilize the prevailing domain restrictions along with some common heuristics for enhancing the precision during matching. For illustration, heuristics is an investigation where two adjacent nodes are possibly equal. The illustration for domain restriction is if a node A is equivalent to team member and node B is a predecessor of A in the catalog then it is unlikely that B is equal to programmer. These limitations occur commonly and the heuristics are used for manual mapping between relations. The technique is based on repose catalogs a strong method employed widely in image processing and modified successfully for solving the problems in matching and taxonomy. The repose catalogs can be successfully employed for handling a variety of heuristics and domain restrictions.

Managing difficult mappings: At last the sealant technique is expanded to design ESealant for identifying difficult mappings between two given catalogs like office maps to branch and working team where ESealant modifies the finding method for effective exploration in such mappings.

Relation mapping: Associations describe the concepts underlying within a domain in terms of observation, elements and associations. These observations provide intended elements of interest within the domain. They are organized into an indexed tree where each node symbolizes a percept and each percept is a hereditary from their parents. Figure 1 shows two examples indexing for a person working for an organization.

Each observation in the index is associated with a collection of designs. For illustration the observation programmer has an occurrence "XYZ" and "ABC" as depicted in fig. 1. Based on the definition of catalog the occurrence of a perception are also due to the occurrences of a predecessor percept. For illustration the percept programmer, team member, developer in Fig. 1 are also the occurrences of designation and people.

Each percept is connected with a set of elements. For illustration the percept programmer in Fig. 1 has an element name, branch and team. The occurrence belongs to a percept with a fixed element values. For illustration, a relation advised by (client, programmer) lists all the pairs of client and programmer such that the earlier is advised by the later. Several languages are prevailing for describing the associations proposed only for semantic web like OWL, DAML+OWL, SHOE and RDF. These language are dissimilar in terms of expressions, the associations they devise allocates identical features.

With the given two relations the relation matching problem is to identify semantic mappings between them. The common way of mapping is one-to-one mapping between the elements like programmer to tem member and office maps to branch. It is noted that mappings between several types of elements are possible like the relation advised by (client, programmer) maps to element advisor of the percept client. Examples of several difficult types of mappings involve the name maps to first and the last name and the office maps to the branch. In common the mapping may be precise as a query which modifies occurrences in one catalog into occurrences of the other. The objective is to identify mappings between the catalogs. This is because relations are the core elements of relations and successful matching of them greatly helps in matching rest of the relations.

MATERIALS AND METHODS

The framework: The allotment estimator takes input as two catalogs C_1 and C_2 together with their data occurrences. Then it relates machine based learning technique for calculating every pair of perception

$$\{X \in C_1, Y \in C_2\}$$

be their probability joint distribution as:

$$P(X,Y), P(X,\bar{Y}), P(\bar{X},Y), P(\bar{X},\bar{Y})$$

Finally, a total of $4|C_1||C_2|$ are calculated where $|C_i|$ is the number of nodes in the catalog C_i. The sealant inputs these two into the resemblance estimator which relates a user provided resemblance function for evaluating the values for each percept pairs.

$$\{X \in C_1, Y \in C_2\}$$

The output from these components is a resemblance matrix between the percepts in the two catalogs. The repose percept module holds the resemblance matrix jointly with domain detailed restrictions and heuristic understanding. The exploration for the mapping configuration, that best convinces the domain restrictions and the common understanding and considered for the noted resemblances (Fig. 2).

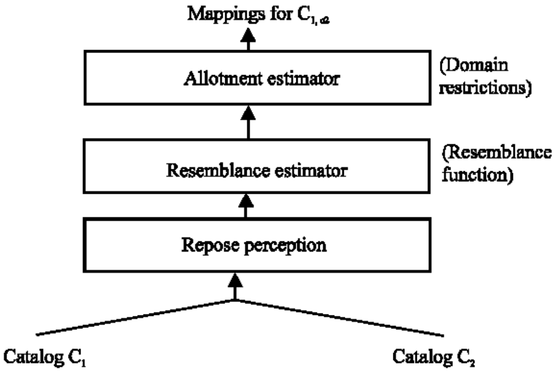

Fig. 2: The framework

Resemblance estimator: For accurate resemblance explanations and to justify how the technique assures inspiration it is necessary to design each perception as a set of occurrences chosen from a several perceptions from the real world. For illustration, for company domain the real world contains all the elements of interests in that world like team member, developer, programmer, client, team leader and so on. The concept programmer is a set of all occurrences in the real world are programmers. For this model the concept of the probability joint distribution between any two percepts X and Y are defined clearly. The distribution consists of:

$$P(X,Y),P(X,\bar{Y}),P(\bar{X},Y),P(\bar{X},\bar{Y}) \quad (1)$$

The term $P(X,\bar{Y})$ is the possibility that a arbitrary selected occurrence from the real world belongs to X but not to Y and is calculated as the fraction of the real world that belongs to X but not to Y.

Several resemblance measures are defined based on the probability joint distribution of the percepts involved.

$$P(X,Y)=\frac{P(X\cap Y)/P(X\cup Y)P(X,Y)}{P(X,Y)+P(X,\bar{Y})+P(\bar{X},Y)} \quad (2)$$

The resemblance measure takes a minimum value 0 when X and Y are not likely and a maximum value 1 when x and Y are the same percepts. Rather attempting to calculate particular resemblance values directly, sealant aims on calculating the probability distributions. It is possible to calculate any of the resemblance measures as a function over the probability joint distributions.

Repose perception: The repose perception is an effective technique for addressing the problem of conveying perceptions to nodes for a given domain sets. The purpose of the technique is that the tag of a node is influenced by the features of the adjacent nodes. The illustration of these features is the tags of the adjacent nodes, the fraction of nodes in the adjacency that convinces a particular conditions and the fact that a particular condition is fulfilled or not.

The repose labeling utilizes this study. The authority of a adjacent node on its tag is measured using formula for the possibilities of each tag as a features of adjacent function. The repose perception allocates primary tags to the nodes based only on the essential properties of the nodes for performing limited iterations. Each iterations uses a formula to modify a tag of a node based on the adjacent features. The process is resumed until the tags are not modified from one iteration to the next until the union condition is attained.

The repose perception serves as a solution to the problem because it is been useful for resemblance matching problems, language processing and hypertext categorization. It is effective and can hold a wide range of conditions.

The repose perception is related to the problem of mapping from catalog C_1 to catalog C_2. The observation of nodes in C_2 as tags and recasting the problems as decisions, the best tag transfers to nodes in C_1 given all the understandings about the domain and the two catalogs.

The purpose is to develop a formula for updating the possibilities that a node takes a tag based on the adjacent features. Let A be a node in catalog C_1 and T be a tag and Δ_U represents the understandings about the domain in particular the tree frameworks of the two catalogs, the set of occurrences and the set of domain restrictions. The following is the probability based on constraints.

$$\begin{aligned}P(A=T\mid\Delta_U)&=\sum_{E_A}P(A=T,E_A\mid\Delta_U)\\&=\sum_{E_A}P(A=T\mid E_A,\Delta_U)P(E_A\mid\Delta_U)\end{aligned} \quad (3)$$

Here, the summation is over all the possible tasks E_x to all nodes other than A in catalog C_1. It is assumed that the nodes tag tasks are not dependent of each other given Δ_U as:

$$P(E_A\mid\Delta_U)=\prod_{(A_i=T_i)\in E_A}P(A_i=T_i\mid\Delta_U) \quad (4)$$

Here, E_A and Δ_U represents the understandings about the adjacency of A. To identify the best mapping for a node X of catalog C_1 all the unions over the catalog C_2 are detailed. The calculation for resemblance with respect to X is learned with the highest resemblance. The number of nodes of C_2 casts the problems with matching by exploring a huge perception. For an effective exploration the search is modified at each stage. The Sealant technique is expanded to design ESealant a system that applies the findings to difficult mappings. The ESealant utilizes the information in data and the catalog frameworks for matching purposes and has not utilized the domain restrictions.

Performance analysis: Figure 3 depicts the matching precision for diverse domains and arrangement of sealant structure. In each domain the precision for matching of two different mapping sequences are depicted from the first catalog to the second and vice versa.

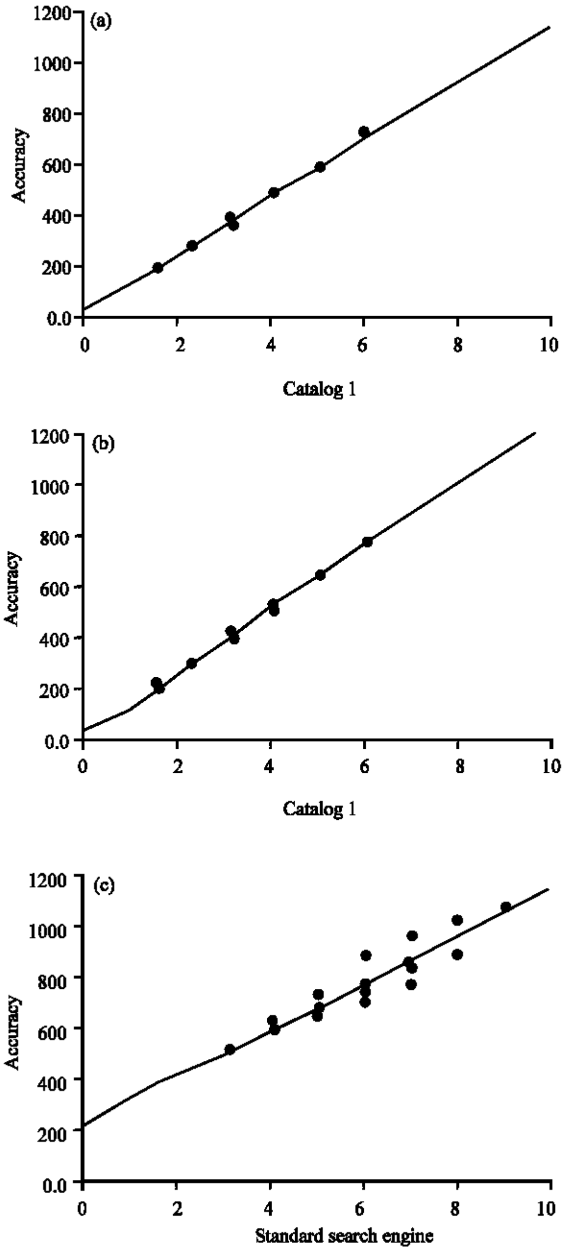

Fig. 3: Matching precision of framework

The results show that the sealant framework attains better performance across all the domains with ranges up to 98%. On comparison, the best matching results of the machine based learning is attained by the perception which is up to 85%. The fact is that the occurrences, for example Y holds identical full names and hence the perception applied for X is also applied to Y where the classification of all the occurrence of Y as X. In case, if the categorization is incorrect which occurs commonly, the perception leads to a poor analysis for the probability joint distribution. The poor performance of the estimator highlights the need for data occurrences and multi state learning in relation matching.

The results clearly depict the value of resemblance and allotment estimator. In most of the cases the estimator modestly enhances the precision and for rest it attains a growth between 8-17%. For all the above cases, the repose perception as added enhances the precision by 4-20% validating that it is able to utilize the domain restrictions and common heuristics. If the repose perception is minimized by 3%, the utilization of other level web applications.

The Sealant framework employs on average only 25- 93 of data occurrences per node in the tree structure. The high precision recommends that the sealant framework works well with a minimal volume of data. The analysis depicts that the repose perception works rapidly and it takes only few seconds for completing iterations. The study reveals that the repose perception can be designed effectively in the relation matching environment. It is possible to implement user suggestions into repose perception process in the form of further domain limitations.

CONCLUSION

Due to the increase in data distribution applications that involves several relations, the design of computerized methods for relation matching serves for their success. A technique is designed for utilizing machine based learning for matching relations. The technique is governed by the sealant technique based on semantic resemblance in terms of probability joint distribution. The purpose of machine based learning particularly multiple-state learning for estimating percepts are explained.

A method named repose perception tags to the relation mapping percepts and viewed that it can be modified effectively for developing different understandings and domain based restrictions for supplementary enhancement matching precision. The framework sealant has been expanded to sealant for identifying difficult mappings between the relations. The extension was planned to hold several complex mappings between the relations including elements and relations.

REFERENCES

Agresti, A., 1990. Categorical Data Analysis. 1st Edn., John Wiley and Sons, Inc., New York, USA.

Berlin, J. and A. Motro, 2002. Database schema matching using machine learning with feature selection. Proceedings of the 14th International Conference, CAiSE 2002 Advanced Information Systems Engineering, May, 27-31, 2002, Springer Berlin Heidelberg, Toronto, Canada, ISBN: 978-3-540-43738-3, pp: 452-466.

Berners-Lee, T., J. Hendler and O. Lassila, 2001. The semantic web. Sci. Am., 284: 34-43.

Brickley, D. and R.V. Guha, 2000. Resource Description Framework (RDF) schema specification 1.0. W3C Candidate Recommendation 27. http://www.w3.org/TR/2000/CR-rdf-schema-20000327/.

Broekstra, J., M. Klein, S. Decker, D. Fensel and F.V. Harmelen *et al*., 2002. Enabling knowledge representation on the web by extending RDF schema. Comput. Netw., 39: 609-634.