Ontological Learning for Analysis of User Preferences

Dhanshri Karbhari[1], Sukanya Naikar[2], Priya Sharma[3], Sharvari Patil[4]

[1] Student, Dept. of Computer Engineering, Universal College of Engineering, Maharashtra, India
[2]Student, Dept. of Computer Engineering, Universal College of Engineering, Maharashtra, India
[3]Student, Dept. of Computer Engineering, Universal College of Engineering, Maharashtra, India
[4]Professor, Dept. of Computer Engineering, Universal College of Engineering, Maharashtra, India

---***---

Abstract - *Hiring is a tedious process for both the candidate as well as the company. Many a times a single wrong hiring can cause millions of losses for the company and a lifetime loss for the candidate as he would be blacklisted and disowned by the whole sector of the companies. With growth of internet and a herd of online job portals that primarily work on a paid growth for profiles, the hiring process becomes worse. People are ready to pay loads of money to job portal sites to get themselves noticed. Recently it has been viewed that many of the profiles on job portals are manipulated or doctored so as it get them noticed. This has lead to breaking down ties with the portals and companies are now moving back to the traditional walk in system. But even this process is extremely hectic, considering the number of applicants that apply for a single post. Due to all this problems there is need for a way to properly screen candidate and offer specialized candidate suitable to the needs of company. This is achieved through the process of ontological learning of the candidate profile. This process helps to easily filter proper candidates that are suitable for a certain job. The ontological profiling is done on the summary of work as well as the live work submitted by the candidate.*

Key Words: **User preferences, Generalization, Sampling, Ontological mapping, Regression.**

1. INTRODUCTION

An onset, information superhighway provided multiple ways of interacting between us, revealing substantial communal network structures, a sensation amplified by internet 2.0 applications. Researchers extracted social networks from emails, mailinglist archives, hyperlink strategy of homepages, co-occurrence of names in documents and from the digital traces created by internet 2.0 consideration usage. Facebook, LinkedIn or Internshala provide immense amounts of structured network data. The emergence of the semantic web approaches attracted to researchers to establish models of a well known online interactions by the agency of ontologies relish FOAF (Friends of Friends), SIOC (Semantically Interlinked Online Communities) or SCOT (Social Semantic Cloud of Tags). This handout starts by the whole of a temporary state of the capability on these enhanced RDF-based representations. We will navigate that the graphs built by the agency of these ontologies have a abundant potential that is not appropriately exploited so far.

An ontology called Semantic SNA is devoted leverage social broadcast and conclude the generally one spontaneous life of an analysis. Finally, we reveal some results based on genuine social disclosure (we extracted thousands of FOAF profiles from users of facebook.com) and urge the advantages, shortcomings and perspectives of a well known an act to appropriately realize semantic social consolidate analysis.

1.1 Ontology for the Semantic Web

The Semantic Web relies severely on the firm ontologies that structure concealed disclosure for the motive of all the options and transportable mechanism understanding. Therefore, the success of the Semantic Web depends in a satisfactory manner on the proliferation of ontologies, which requires expeditious and easy engineering of ontologies and avoidance of a development acquisition bottleneck.

1.2 Ontology Learning process steps

First, at this moment ontologies are imported and reused by merging prompt structures or defining mapping rules between critical structures and the ontology expected established. For instance, describe how ontological structures contained in Cycle are secondhand in term to assist the interpretation of a domain-specific ontology.

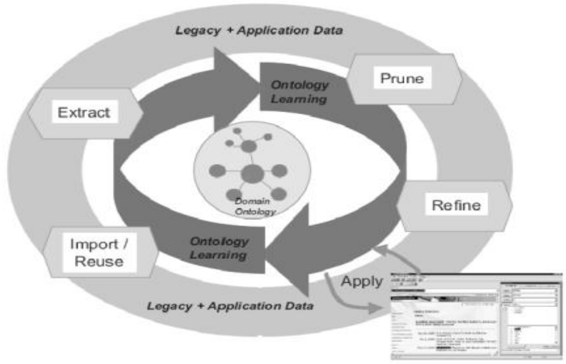

Fig-1: Learning process steps

Second, in the ontology parentage phase masterpiece parts of the intend ontology are modeled mutually learning contend feeding from internet documents. Third, this serrated outline of the target ontology needs impending pruned in censure to better mediate the ontology to its dawn purpose. Fourth, ontology leniency profits from the supposing domain ontology, yet completes the ontology at a fine granularity (also in analyze to extraction). Fifth, the dawn target consideration serves as analyze for validating the resulting ontology. Finally, one commit revolve still in this bi bike, e.g. for including new domains directed toward the constructed ontology or for maintaining and updating its scope.

1.3 Artificial Neural Networks

One description of unite sees the nodes as 'artificial neurons'. These are called artificial neural networks (ANNs). Natural neurons sip signals on synapses entrenched on the dendrites or membrane of the neuron. When the signals confirmed are fortunate enough (surpass a certain threshold), the neuron is activated and emits all hail though the axon.The complication of genuine neurons is highly inattentive when modelling artificial neurons. These basically comprise inputs (like synapses), which are assorted by weights (strength of the respective signals), and earlier computed by a mathematical employment which determines the activation of the neuron. Another work (which make out be the identity) computes the produce of the exaggerated neuron (sometimes in dependence of a certain threshold).

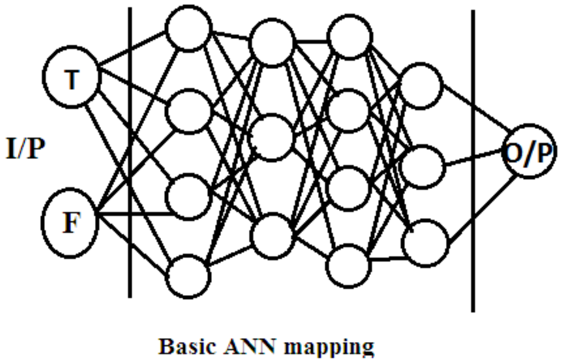

Fig-2: ANN'S Mapping

The higher a load of an exaggerated neuron is, the stronger the input which is multiplied by it will be. By adjusting the weights of an artificial neuron we can bring in the output we desire for stead inputs.

2. PROBLEM STATEMENT

The ontological learning approach has paved its way in the field of analytics and major tech giants are using ontological mapping techniques to map the requirements of the user and hit him with the right ads.

The global trends suggest it is extremely hard to find the right person for the right job. A lot of online job portals operate across all the countries and offer lucrative offers to candidates without actually screening the candidates ability to do the job. In most of the cases the filtering is done based on educational qualification of the candidate. But the companies require a lot more than what meets the eye. The companies require the following things that are supposed to build in the proposed system:

1) Easy screening of candidates.
2) Filtering candidate by individual or a group of skills.
3) Comparative analysis of two candidate profiles.
4) Rating for subjective analysis of the candidate.
5) Analysis of candidate's review by his colleagues from previous job.

3. LITERATURE SURVEY

Table -1: Literature Survey

Authors	Title	Key Issues	Contribution
Carlos Gershenson	Artificial Neural Networks for Beginners	Introduction to ANN	Exploiting SW for tackling new and challenging research.
Peter Mika	Ontologies are us: A unified model of social networks and semantics	Leading to a tripartite model of actors, concepts and instances.	This model is built on an implicit realization of emergent semantics. Inspired by social tagging mechanisms, concept Andinstance associations.
Li Ding, Tim Finin, Anupam Joshi	Analyzing Social Networks on the Semantic Web	Artificial Neural Networks (ANNs), Back propagation algorithm	It would be useful for people who are just curious about what are ANNs and its Scope.
Alexander Maedche and Steffen Staab	Ontology Learning for the Semantic Web	Social relations	Provides a powerful distributed mechanism to represent and publish social network information.
Xavier Polanco, Claire François, Jean-Charles Lamirel	Using Artificial Neural Networks For Mapping Of Science And Technology: A Multi Self-Organizing Maps Approach	Clustering, cartography, and hypertext generation	Naming the clusters, the map division into logical areas, and the map Generalization mechanism.
Lucas Drumond and Rosario Girardi	A Survey of Ontology Learning Procedures	Difficult and time consuming	Knowledge about a specific domain and its representation in an ontology like structure
Xing Jiang, Ah-Hwee Tan	Learning and inferencing in user ontology for personalized Semantic Web search	Capture a user's interests in the working domain, Traditional methods	User ontology, providing personalized information service in the Semantic Web.

4. PROPOSED SYSTEM

The Proposed System uses the NLP and knowledge extraction as a basis for generating the ontological profile of an individual. The ontological profile formed here is synonymous to a directed graph where all the ontologies are bound together as nodes under the person's node. This model helps to link every individual under some common pretext or subject of expertise. Pruning occurs at every stage of the graph to seperate individual data from global data. The proposed system uses the analytical composition framing to build up data from scratch. The layers build upon each other in the sense that results of tasks at lower layers typically serve as input for the higher layers, similar to the OSI model in computer networks. For example, in order to extract relations between concepts, consider the overall hierarchy to identify the exact level of generalization for the domain and range of the relation. The two bottom layers of the layered structure correspond to the lexical level of ontology learning.

5. METHODOLOGY

The proposed system is to be implemented using the spiral model in conjunction with verification and validation at every stage of development. The verification and validation at every stage enables us to rectify any errors that we might encounter in any given stage. The project is split into following modules:

5.1 Knowledge Extraction Module

The knowledge extraction module is one of the most important module in our system. This module provides us with the information of the users and their activities over our website. The knowledge extraction process is done using serial mapping of events of clickstream data and logs that are generated. The ontologies created here will be stored in form of processed indices.

5.2 Natural Language Processing

The NLP here is done using Naive Bayes. We opt for Naive Bayes as we do not want to do the sentimental analysis but the objective analysis of the data that we have collected through the comments section. The comments over the users profile will be considered in the ontological analysis of the user. The ontologies formed by the personality profile with comments are combined to form overall analysis.

5.3 Profile Filtering

The profile filtering is deployed at both the sides. The user side and company side logins both will be having a mode to filter profiles based on the set of requirements.

6. SYSTEM ARCHITECTURE AND WORKFLOW

6.1 Term Extraction

The task at the lexical layers is to bifurcate terms and rearrange these into groups of synonymous words. A simple technique for the extraction of relevant terms that may indicate concepts is counting the frequencies of the terms in a given set of (processed at linguistic level) documents, the corpus D. In general this approach is based on the assumption that a frequent term in a set of domain-specific texts indicates the occurrence of a relevant concept. Research in information retrieval has shown that there are more effective methods of term weighing than simple counting of frequencies.

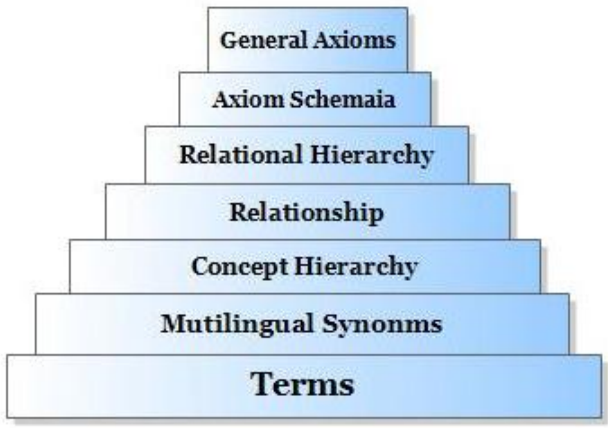

Fig-3: Level of extraction

6.2 Synonym Extraction

In order to extract synonyms, most approaches rely on the distributional hypothesis claiming that words are semantically similar to the extent to which they share syntactic contexts. The pointwise mutual information of two events x and y is defined as:

$$PMI(x,y) := \log_2 \frac{P(x,y)}{P(x)P(y)}$$

where P(x, y) is the probability for a joint occurrence of x and y and P(x) is the probability for the event x. The PMI is thus in essence the (logarithmic) ratio of the joint probability and the probability under the assumption of independence. In fact, if $P(x, y) \leq P(x)P(y)$, we will have a negative (or zero) value for the PMI, while in case $P(x, y) > P(x)P(y)$, we will have a positive PMI value. The PMI can be calculated using indexing hits and counting hits as follows:

$$PMI_{web}(x,y) := \log_2 \frac{Hits(x\ AND\ y)\ MaxPages}{Hits(x)\ Hits(y)}$$

Where
MaxPages is an approximation for the maximum number of English web pages. This measure can thus be used to calculate the statistical dependence of two words on the Web. If they are highly dependent, we can assume they are synonyms or at least highly semantically related.

6.3 Clustering

Clustering can be defined as the process of organizing objects into groups whose members are similar in some way based on a certain representation, typically in the form of vectors or iterative formed data structures. In general, there are three major styles of clustering:

6.3.1 Agglomerative

In the agglomerative type of clustering, in the initialization phase, every term is defined to constitute a cluster of its own. In the growing or the growth phase, the larger clusters are iteratively generated by merging the most similar/least dissimilar ones until some stopping criteria is reached.

6.3.2 Divisive

In the initialization phase, the set of all terms constitutes a cluster. In the refinement phase, smaller clusters are (iteratively) generated by splitting the largest cluster or the least homogeneous cluster into several subclusters. Both agglomerative and divisive clustering techniques are used to produce hierarchical descriptions of terms.

6.3.3 Conceptual

Conceptual clustering builds a lattice of terms by investigating the exact overlap of descriptive attributes between two represented terms. In the worst case scenario, the complexity of the resultant concept lattice is exponential in n. Here's how concisely the system will map all the resources:

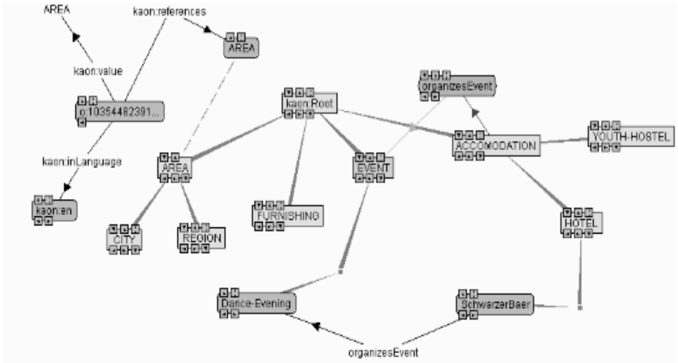

Fig-2: Mapping of the Resources

7. FUTURE WORK

We proposed a semantic SNA stack to better exploit the rich representation of online Professional social networks and leverage social data with SNA features. Our perspectives include the adaptation of other algorithms in particular for community detection, and new semantic algorithms based on adaptation of classical SNA definitions. We will consequently extend SNA to semantically describe these new features and also help their querying extraction with inference rules on top of social data. Future work includes the development of iterative algorithms and methods to manage evolutions of ever-changing networks.

8. CONCLUSIONS

In this paper, we have proposed an extensible fine-grained online Professionals social network access control model based on semantic web tools. In addition, we propose authorization, administration and filtering policies that are modeled using OWL and SWRL. The architecture of a framework in support of this model has also been presented. Further, we have implemented a version of this framework and presented experimental results for the length of time access control can be evaluated using this scheme. Further work could be conducted in the area of determining a minimal set of access policies that could be used in evaluating access requests in a further attempt to increase the efficiency of these requests. Additionally, we have shown that existing Professional social networks need some form of reasonable data partitioning in order for semantic inference of their access control to be reasonable in its speed and memory requirements, due to constraints on the memory available to perform inference. Additionally, further work can be used in determining the best method of representing the individual information of a person in a Professional social network to determine if a hybrid semantic or relational approach or a pure approach offers the best overall system.

REFERENCES

[1] Peter Mika, Ontologies are us: A unied model of social networks and semantics, Vrije Universiteit, Amsterdam 1081HV Amsterdam, (Feb 2007)5-15.
[2] Li Ding, Tim Finin, Anupam Joshi, Analyzing Social Networks on the Semantic Web, University of Maryland Baltimore County, (Dec 2004).
[3] Alexander Maedche1 and Steffen Staab, Ontology Learning, FZI Research Center for Information Technologies, University of Karlsruhe, Germany.
[4] Alexander Maedchel , Steffen Staab, Ontology Learning for the Semantic Web, Institute AIFB, D-76128 Karlsruhe, Germany and Ontoprise GmbH, Haid-un Neu- Strasse 7, 76131 Karlsruhe, German Word Count: 5541.
[5] Carlos Gershenson, Artificial Neural Networks for Beginners.
[6] Xavier Polanco, Claire François, Jean-Charles Lamirel, Using Artificial Neural Networks For Mapping Of Science And Technology: A Multi Self Organizing Maps Approach, Uri/Inist-Cnrs, 2,allée du Parc de Brabois – 54514.
[7] Sampson, D. G., Lytras, M. D., Wagner, G., & Diaz, P. (2004). Ontologies and the Semantic Web for E-learning. Educational Technology & Society, 7 (4), 26-28.