

Use of Bayesian Network in Information Extraction from Unstructured Data Sources

Quratulain N. Rajput, Sajjad Haider

Abstract—This paper applies Bayesian Networks to support information extraction from unstructured, ungrammatical, and incoherent data sources for semantic annotation. A tool has been developed that combines ontologies, machine learning, and information extraction and probabilistic reasoning techniques to support the extraction process. Data acquisition is performed with the aid of knowledge specified in the form of ontology. Due to the variable size of information available on different data sources, it is often the case that the extracted data contains missing values for certain variables of interest. It is desirable in such situations to predict the missing values. The methodology, presented in this paper, first learns a Bayesian network from the training data and then uses it to predict missing data and to resolve conflicts. Experiments have been conducted to analyze the performance of the presented methodology. The results look promising as the methodology achieves high degree of precision and recall for information extraction and reasonably good accuracy for predicting missing values.

Keywords—Information Extraction, Bayesian Network, ontology, Machine Learning

I. INTRODUCTION

THE increase in the volume of information available over the World Wide Web (WWW) has made it extremely difficult for users to find and utilize information in an efficient manner. One of the main reasons for this inefficiency is the inability of the existing web based information storage and presentation systems to add semantics to the web data [2]. During the last few years, however, semantic web technologies [25, 27] have emerged as a much needed platform for future web development. Semantic web is an extension of the current web in which information is given well-defined meaning, thus making it possible for machines to understand web content [19]. It consists of elements such as RDF/XML, RDF Schema, and OWL which facilitate both website developers and users in expressing formal description of concepts and their relationships [1].

One of the main challenges in fully realizing the goal of semantic web is the semantic annotation of the existing web data. Automated tools are required that can extract information and assign semantics to it. Much of the research

in semantic annotation has been focused on finding relevant data from existing content using information extraction (IE) techniques [17, 20]. Many tools, based on IE techniques, have been reported in the literature. Laender et. Al [13] provides an extensive survey of such tools. Another important category of tools is based on ontologies [9]. These tools support automatic and semiautomatic annotation using domain specific ontologies. The ontologies describe data of interest, their relationship, lexical appearance, and context keywords. Some of the important ontology-based tools for semantic annotation are BYU [6-8], S-Cream [10], MnM [26], iASA [24], and ontoX [28-29].

This paper also presents a methodology that integrates ontology, information extraction techniques, and Bayesian Networks (BN) to extract information from structured/unstructured/ungrammatical data sources and to annotate it with semantic information [21]. The work is an extension of our other reported work describing a tool [21], named OWIE (Ontology based Web Information Extraction). The tool combines ontology and information extraction techniques to support the annotation process. It was observed, however, that to extract information from unstructured and incoherent data sources, one has to deal with variable size of information available at different sources within a single domain. For instance, if one wants to find used cell phones from two different web sites, such as shopping.yahoo.com and ebays.com, then the person may find one website containing very detailed information (brand, model, camera, condition, battery, display, weight, etc.) while the other containing very few data elements (brand and model). Most of the IE systems, including OWIE, when confronted with such situation, store the unavailable data elements as missing. However, it is possible in many situations that the data is missing due to some causal reason. In other words the data is missing because the generator thought it could easily be predicted from other non-missing values. If the end-user of the data is also familiar with such assumptions then no ambiguity arises. However, such assumptions are not typically known to naïve users and they intend to ignore records having missing values. Secondly, to make the data machine readable, these assumptions need to be made explicit. Furthermore, it is typically the case that the relationship between missing and non-missing values is not deterministically causal instead a strong probabilistic relationship exists among them. Thus, it is highly desirable in situations like these, when there is a strong correlation among the missing values and other non-missing values, to accurately predict the missing data to better guide a user during her decision making process. The second

Q.N. Rajput is with the Faculty of Computer Science, Institute of Business Administration, Karachi, Pakistan (e-mail: quratulain.rajput@gmail.com phone: (92-21)111677677; fax: (92-21)9215528).

S. Haider is with the Faculty of Computer Science, Institute of Business Administration, Karachi, Pakistan (e-mail: sajjad.haider@khi.iba.edu.pk).

important limitation of IEs is their inability to resolve homonyms, hypernym, hyponymy, and synonyms. In some cases, context words can aid in resolving this issue but the situation becomes more complicated when the relevant context words are not available in the text. This paper applies probabilistic reasoning techniques, commonly known as Bayesian Networks, to address these problems. An enhanced version of OWIE, named E-OWIE, is presented and is applied on a real data set with many missing values.

The rest of the paper is organized as follows. Section II briefly describes Bayesian Networks while Section III explains E-OWIE (Enhanced Ontology-based Web Information Extraction) and how it applies Bayesian networks to predict the unavailable/missing values and to resolve conflicts. Section IV discussed the performance of E-OWIE on a case study. Finally, section V concludes the paper and provides the future research directions.

II. BAYESIAN NETWORKS

During the last two decades, Bayesian Networks (BNs) [4,5,18] have emerged as the most popular tool for modeling and reasoning in uncertain domains. They have been successfully applied in several domains including medical diagnosis, terrorism, forecasting, information fusion, system troubleshooting, etc. Mittal and Kassim [30] and Pourret et al. [31] provide an extensive list of areas where BNs have been applied. A BN is a graphical representation of probability distributions. It consists of two components. The first is a directed acyclic graph in which each node represents a random variable, while the set of arcs connecting pairs of nodes represents certain conditional independence properties. This component captures the *structure* of the probability distribution. The second component is a collection of *parameters* that describe the conditional probability of each variable given its parent in the graph. Together, these two components represent a unique probability distribution [18]. Figure 1 shows a sample BN for a medical diagnosis problem. There are five random variables (nodes). The links connecting different variables show the direct dependency among these variables. The values attached to the nodes represent the corresponding prior and conditional probabilities. In general, the joint probability distribution of variables (X_1, X_2, \dots, X_n) in a BN is computed using these local distributions thus easing the process of knowledge elicitation. Mathematically,

$$P(X_1, X_2, \dots, X_n) = \prod P(X_i | pa(X_i))$$

where $pa(X_i)$ represents the set of parents of variable X_i .

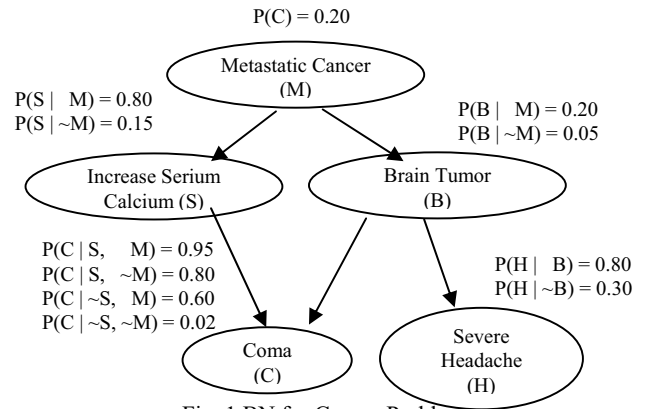


Fig. 1 BN for Cancer Problem

Traditionally the practice has been to build a BN by acquiring knowledge from domain experts. However, if data is available a BN can be learned from that data. In the past few years, a lot of work has been done for learning both the structure and the parameters of a BN from a data set [3, 11, 12, 14, 15, 16, 22, 23]. In all such efforts, the uncertainty about the structure of a BN is represented as a probability distribution over all possible structures. Prior distributions are assigned to each structure and to the local distributions given that particular structure. Then Bayes rule is applied to obtain the posterior distribution for structures and parameters given data D . Mathematically,

$$P(S_i | D) = \frac{P(D | S_i)P(S_i)}{P(D)}$$

Where $P(D)$ is a normalization constant that does not depend upon any structure. $P(D | S_i)$ is the marginal likelihood of function, $P(S_i | D)$ is the posterior distribution of the structure given the data, and $P(S_i)$ is the prior probability of the structure. An excellent overview of the BN learning process is given by Heckerman [32].

III. E-OWIE: ENHANCED ONTOLOGY-BASED WEB INFORMATION EXTRACTION

This section explains the working of E-OWIE, which is an enhanced version of OWIE. A brief description of OWIE is described next while readers interested in its detailed working should refer to [21]. The OWIE tool was originally developed to extract information from unstructured and ungrammatical data sources such as used laptop ads posted by various users at craigslist website¹. It finds and extracts relevant information with the help of a pre-defined ontology. The process starts with retrieving links of data of interest from explicitly provided URL(s). In an iterative manner, each link is explored which contain ad description posted by different users. The extraction application module takes domain ontology and ad description as input and perform extraction using rules by exploiting knowledge stored in ontology as shown in Fig. 2.

¹ <http://www.craigslist.com>

This knowledge is stored in the form of concepts, relationships among concepts, data type properties, and context words. The context words are stored in the comment section associated with each concept and data type properties. The rules are defined as regular expression to describe the appearance of the value to be extracted.

The extracted data generally contains some missing value because of the unavailability of information. However, in many cases there are missing data elements which, if predicted, can increase the quality of the semantic annotation process. The integration of BN into the OWIE tool aims to address this issue. Furthermore, the integration can aid in resolving different types of conflicts present in the extracted data. The enhanced OWIE which makes use of BN for prediction and conflict resolution is termed as E-OWIE. The graphical description of the complete process of E-OWIE is shown in Fig. 2. The process is mainly divided into three different modules:

- 1) Ontology-based information extraction
- 2) Bayesian network learning
- 3) Missing values prediction.

The primary focus of this paper is on the remaining two modules. The first module is the same as defined in our other work [27] and has been briefly described above. The process starts with instance extraction module retrieving information of interest from explicitly provide URL(s). In an iterative manner, each link is explored and required information is extracted using knowledge define in ontology. The values are then stored in a file. If there are missing values in the data then these values are passed to the next stage as missing. Once all the links are extracted by the extraction application module, the data set (with both missing and non-missing) information is passed to the Bayesian network learning module. This stage can be considered as the training phase and is executed only once for learning the BN. After cleaning and processing the data, the module learns the structure and parameters of a BN and the resultant network is used during the missing value prediction and conflict resolution processes performed by the third module. This third stage is executed during the actual implementation of E-OWIE. Once the data passes through the third module, it is annotated using OWL/RDF. Tables I and II present the main steps involved in the second and third modules while a detailed discussion both modules follows next using craigslist website as a case study.

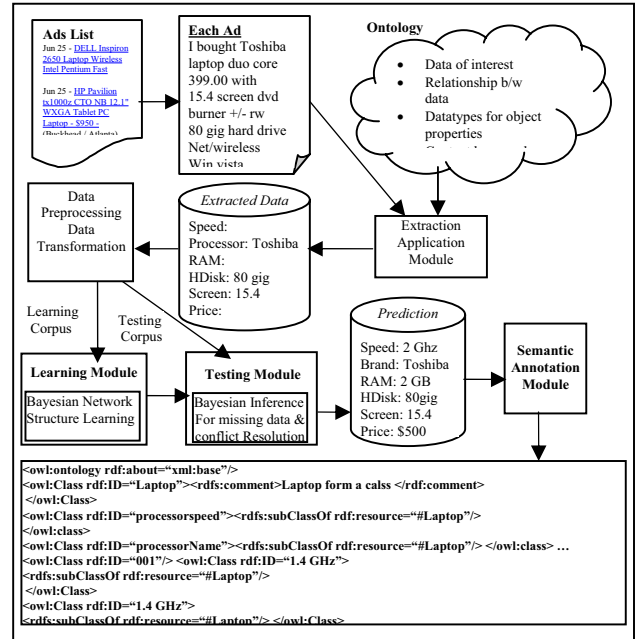


Fig. 2 Process Model of E-OWIE Tool

TABLE I
 LEARNING ALGORITHM

```

BEGIN
Step 1: Read each record from file
Step 2: For each record
    i. If the record is incomplete
        Prune it and move to the next record
    Else
        Verify extracted data for correction
    ii. Discretized continuous data and store in a file
Step 3: Learn the structure of Bayesian Networks using data
obtained in Step 2.
END
    
```

TABLE II
 TESTING ALGORITHM

```

BEGIN
Step 1: Read each Record from testing corpus
Step 2: For each Record
    a. Discretized continuous data and store in a file F
Step 3: For each record in file F
    a. Consider available data in a record as hard core
evidence
    b. Predict the missing value of a record using
learned model of Bayesian Networks
END
    
```

1) BAYESIAN NETWORK LEARNIGN MODULE

The purpose of this module is to learn the structure and parameters of a Bayesian network using training data. Once learned, the BN aids in predicting missing values and resolving conflicts in ads. The module has two main steps: (a) data cleaning and pre-processing and (b) learning.

a) Data Cleaning and Pre-Processing and Cleaning

The module gets the extracted data from the information extraction module. For instance, for the craigslist example, the module gets the 6-tuple data as shown in Table III. The elements of the 6-tuple include Speed, Price, Brand, Screen, RAM, and HDisk. It should be noted that the table shows only few records due to space limitation. The original data set selected for the case study contains 2000 records. In general, the actual size of data required for accurately learning a BN is quite large (in thousands or even more) and depends upon many factors including the number of attributes in a problem domain and how many distinct states these attributes have.

To learn a BN, E-OWIE ignores all the missing records during this learning phase. For example, if E-OWIE has to learn from the data set of Table III, it retains only rows 3, 4, and 7 during this stage. When executed on 2000 records in our craigslist case study, this step retains 548 records having no-missing values.

TABLE III

INFORMATION EXTRACTION FROM CRAIGSLIST SITE

S.No	Speed (GHz)	Price (\$)	Brand	Screen (inch)	RAM (GB)	Hdisk (GB)
1	1.4 Ghz	\$250		14 inch	768 MB	40GB
2	2.4GHz				1GB	60GB
3	1.5 Ghz	\$400	Dell	14"	2MB	120GB
4	2.4GHz	\$2,100	Dell	24 inch	4GB	1024GB
5		\$70		17"		
6	2.2ghz	\$275	dell			20GB
7	1.4GHz	\$750	Dell	15.4 inch	2GB	250GB
8		\$225				
9	1.59 GHZ	\$450			896 MB	44GB

In the next stage, all the continuous data columns are discretized. For the craigslist example, the following intervals are defined to categorize each attribute:

- Speed= {1-1.5, >1.5-1.75, >1.75-2, > 2-2.25, >2.25-2.75, 2.75-4}, in GHz.
- Price= {100-300, >300 to 500, >500 to 700, >700 to 900, >900 to 1099, >1100 to 2000}, in \$
- Brand= {HP, DELL, SONY, IBM, TOSHIBA}
- Screen= {10-12, >12-14, >14-15, >15-16, >16-17, >17-21}, in inch
- RAM= {0-1, >1-2, >2-3, >3-4, >4-5}, in GB
- HDisk= {10-60, >60-100, >100-200, >200-300, >300-500, >500-1000}, in GB

b) Learning

Once the data is cleaned from missing values and all the continuous attributes are discretized, the next task is to capture the probabilistic relationships that exist among the attributes by learning the structure and parameters of a BN. It is important to note that the structure learning of a BN is an intractable task and thus no polynomial time algorithm exists. Most of the existing schemes are based on greedy approaches

or computational intelligence techniques (such as evolutionary algorithms, particle swarm optimization, etc.). E-OWIE accomplishes this task with the aid of BN-PowerConstructor². The tool learns the structure and parameters of a BN using heuristics. It allows the incorporation of certain types of causal/non-causal constraints during the structure learning process. For the craigslist case study, the 500+ records were provided to BN-PowerConstructor and the tool learned the structure and parameters of a BN that captures the underlying probabilistic relationships among the attributes. The BN is shown in Fig. 3 the graphical interface of BN-PowerConstructor only displays the structure of the learned BN. To study the parameters of the learned model and the associated marginal probabilities, the BN is analyzed using GeNIe³ shown in Fig. 4.

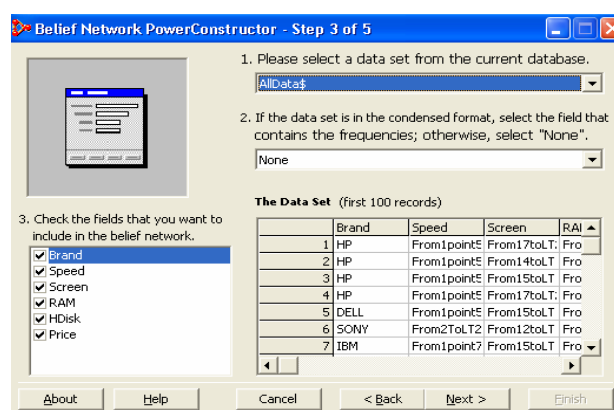


Fig. 3 Structure Learning with Belief Network Power Constructor

2) MISSING INFORMATION PREDICTION MODULE

Once a BN is successfully learned during the training phase, it is ready to be used in the implementation phase of E-OWIE for missing value prediction and conflict resolution. During this stage, a user types a query to search for a specific piece of information. The information extraction phase first retrieves links from the pre-specified websites and then extracts the relevant information and passes it to the data cleaning and pre-processing stages of the BN module. If there are missing and/or conflicting values in the resultant data set, the data pre-processing phase passes it to missing information prediction module. The module uses the Bayesian inference mechanism to predict missing values and resolve conflicts. All the non-missing/non-conflicting values are considered as hard evidence and are entered as such in the learned BN. As a result of the hard evidence, the belief updating algorithm updates the probabilities of missing/conflicting values. If the marginal probabilities of any of their states are above the threshold probability then E-OWIE considers the missing/conflicting data to belong to that state.

² BN-Power Constructor can be downloaded from <http://www.cs.ualberta.ca/~jcheng/bnpc.htm>.

³ GeNIe is a Bayesian network editor developed by the Decision Systems Laboratory, University of Pittsburgh. The software can be downloaded from <http://genie.sis.pitt.edu/>

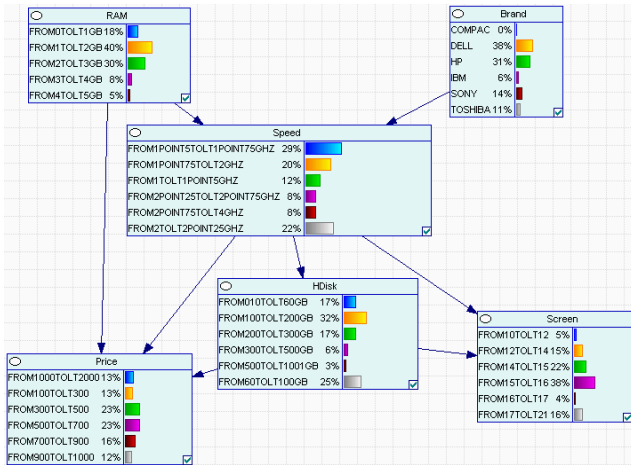


Fig. 4 Learned Model for Prediction

TABLE V
 INSTANCE VALUES TRANSFORMED IN RANGE

S.No	Speed (GHz)	Price (\$)	Brand	Screen (inch)	RAM (GB)	Hdisk (GB)
1	2.75 to 4	500 to 700	HP	17 to 21	0 to 1	200 to 299
2	2 to 2.25	900 to 2000	DELL	17 to 21	2 to 3	200 to 299
3	1.5 to 1.75	1100 to 2000	IBM	12 to 14	3 to 4	200 to 299
4	2 to 2.25	700 to 900	IBM	15 to 16	2 to 3	60 to 99
5	1.75 to 2	700 to 900	HP	14 to 15	3 to 4	200 to 299

IV. EXPERIMENTS

To test the performance of E-OWIE, we have selected 76 ads (test data) from the craigslist website. These ads are distinct from the training ads corpus used in the previous section to learn the BN. The ads contain values for all six attributes. The values are manually verified to make sure that they contain no incorrect data item. Out of these 76 ads, 10 are listed in Table IV. As discussed in the previous section, these values are first discretized according to the scheme mentioned above. Table V shows the transformed data of Table IV. The experiment is conducted in the following way. Out of all the six data items in a record, each is considered missing once and is predicted using the remaining 5 values. For instance, we assume that the Speed of the first record in Table V (first row) is unknown and predict it using the other five values (Price=500-700, Brand=HP, etc.). If the prediction matches with the actual known value then the prediction is right, otherwise it is wrong. The process is repeated for other five variables in the record, i.e., consider the Price data in row 1 of Table V as missing and predict it using Speed, Brand, etc. and so on. Thus, for our sample data set of 76 records, the process is repeated 76 x 6 times. The predicted values of Table V data are given in Table VI.

TABLE VI
 TESTED RESULTS FOR PREDICTION

Sno	Speed (GHz)	Price (\$)	Brand	Screen (inch)	Ram (GB)	Hdisk (GB)
1	1.75 to 2	100 to 300	SONY	17 to 21	1 to 2	200 to 299
2	2 to 2.25	700 to 900	HP	12 to 14	2 to 3	200 to 299
3	2 to 2.25	500 to 700	HP	15 to 16	0 to 1	100 to 199
4	2 to 2.25	500 to 700	HP	14 to 15	2 to 3	100 to 199
5	2 to 2.25	900 to 2000	DELL	15 to 16	2 to 3	60 to 99
6	2 to 2.25	901 to 2000	DELL	15 to 16	3 to 4	100 to 199
7	1 to 1.5	100 to 300	HP	14 to 15	1 to 2	100 to 199
8	2 to 2.25	500 to 700	HP	15 to 16	2 to 3	010 to 60
9	1.5 to 1.75	700 to 900	HP	15 to 16	3 to 4	100 to 199
10	1.75 to 2	500 to 700	DELL	15 to 16	2 to 3	100 to 199

TABLE IV
 INFORMATION EXTRACTED FOR TESTING

S.No	Speed (GHz)	Price (\$)	Brand	Screen (inch)	RAM (GB)	Hdisk (GB)
1	2.93	550	HP	17	0.5	200
2	2.2	1000	DELL	17	2	250
3	1.5	1399	IBM	12.1	3	200
4	2	725	IBM	15.4	2	80
5	1.86	795	HP	14.1	3	250
6	1.9	900	HP	12.1	3	250
7	1.9GB	300	IBM	15	0.5	30
8	1.6	799	DELL	14.1	3	160
9	2	650	TOSHIBA	17	2	160
10	1.9	699	HP	15.4	3	160

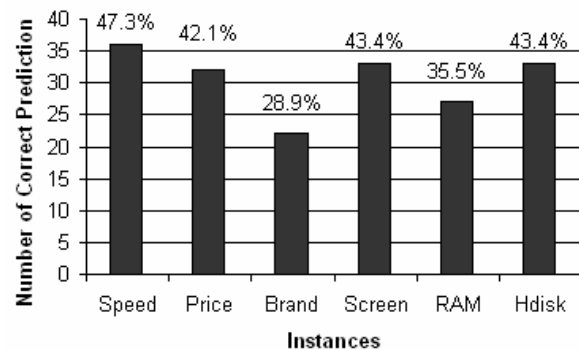


Fig. 5 Prediction Result of E-OWIE Tool

TABLE VII
 PREDICTION RESULT

Sno	Speed (GHz)	Price(\$)	Brand	Screen (inch)	Ram (GB)	Hdrive (GB)
1	0	0	0	1	0	1
2	1	0	0	0	1	1
3	0	0	0	0	0	0
4	1	0	0	0	1	0
5	0	0	0	0	0	0
6	0	1	0	0	1	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0
9	0	0	0	0	0	1
10	1	1	0	1	0	1

	47.3684	42.1052	28.9473	43.4210	35.5263	43.4210
%	2	6	7	5	2	5

The accuracy of BN-based E-OWIE predictions are measured by recording the percentage of correct prediction. For the current example of craigslist ads, the accuracy can be measured by comparing the predictions of Table VI to the original value of Table V. If the data in a particular cell is same in both Tables V and VI, it implies the correctness of the prediction. Table VII shows such comparison. A value of 1 indicates a correct prediction while a value of 0 indicates an incorrect prediction. The last row shows the accuracy of E-OWIE for each attribute in the data set. A bar chart based on these accuracy estimates is shown in Fig. 5. The combine prediction accuracy is 40.1% for all 76 ads. It was observed that the data attributes selected in the laptop case study were not highly correlated. This had resulted in an average performance of E-OWIE. It is strongly believed that if a different domain having a strong correlation among data elements is selected, E-OWIE could achieve a much higher accuracy.

V. CONCLUSION

The paper presented a methodology based on ontology, Bayesian Networks, and information extraction techniques to extract data from unstructured and ungrammatical sources and to add semantic annotation to it. The major contribution of this work is the incorporation of Bayesian Networks to predict the missing data and to resolve conflicts that arise during the extraction process. A tool, named E-OWIE, has been developed to support the presented methodology. The performance of E-OWIE was tested on a real data set originating from a website that deals in selling/purchasing of used laptops. The results showed that E-OWIE performed reasonably well while predicting the missing data. The limitation in the performance is due to the fact that the data attributes selected in the laptop case study are not highly correlated. It is strongly believed that when applied to a different domain, having a strong correlation among data elements, E-OWIE could achieve a much better accuracy. The application of E-OWIE to different domains is part of our future research directions. Secondly, the current work was

focused on extracting data from a single website. In the future, the focus would be on extracting information from multiple websites dealing in similar kind of business and then applying E-OWIE to the extracted data.

REFERENCES

- [1] Antoniou, G., Harmelen, F.V.: A Semantic Web Primer. 2nd Edition. MIT Press (2004)
- [2] Buneman P., Semistructured data. In Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems. Arizona, United States (1997) 117-121
- [3] Buntine, W. L. (1996), "A Guide to the Literature on Learning Probabilistic Networks from Data," *IEEE Transactions on Knowledge and Data Engineering*, 8, pp. 195-210
- [4] Chariak, E. (1991), "Bayesian Network without Tears," *AI Magazine*, Winter
- [5] Cooper, G. F. and E. Herskovits (1992). "A Bayesian Method for the Induction of Probabilistic Networks from Data." *Machine Learning* 9: 309-347.
- [6] Embley D.W., Ding Y., Liddle S. W., and Vickers M.: Automatic Creation And Simplified Querying Of Semantic Web Content. In Proceedings of First Asian Semantic Conference (ASWC), Beijing China (2006)
- [7] Embley, D.W., Campbell, D.M., Jiang, Y.S., Liddle, S.W., Lonsdale, D.W., Ng, Y.k., Smith, R.D.: Conceptual-Model-Based Data Extraction from Multiple-Record Web Pages. *Journal of Data and Knowledge Engineering*, Vol.31(3), (1999) 227-251
- [8] Embley, D.W., Tao, C., Liddle, S.W.: Automating the Extraction of Data from HTML Tables with Unknown Structure. *Journal of Data & Knowledge Engineering*. Vol. 54(1), (2005) 3-28
- [9] Gruber, T. R. A Translation Approach to Portable Ontology Specifications. 1993
- [10] Handschuh, S., Staab, S., Ciravegna, F.: S-CREAM Semi-automatic CREAtion of Metadata. In Proceedings of 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW), Siguenza Spain (2002)
- [11] Hrycej, T. (1990), "Gibbs Sampling in Bayesian Networks," *Artificial Intelligence*, pg. 351-363
- [12] Jensen, F. V. (2001), *Bayesian Networks and Decision Graphs*, Springer-Verlag.
- [13] Laender, A.H.F., Ribeiro-Neto, B.A., da Silva A.S., Teixeira J.S.: A Brief Survey of Web Data Extraction Tools. In ACM SIGMOD Record, Vol. 31(2) (2002) 84-93
- [14] Laskey K.B., Myers J.W., DeJong K. A., Learning Bayesian Network from Incomplete data using Evolutionary Algorithms. In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence. George Mason University (1999).
- [15] Lauritzen, S. L. (1995), "The EM Algorithm for Graphical Association Models with Missing Data", *Computational Statistics and Data Analysis*, 19, pp. 191-201.
- [16] Neapolitan, R. E. (2003), *Learning Bayesian Networks*, Prentice Hall, 2003.
- [17] Partee, Barbara H., Alice ter Meulen and Robert E. Wall (1990). *Mathematical methods in linguistics*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- [18] Pearl, J. (1987), *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference*, Morgan Kaufmann, 1987.
- [19] Peter and Mika, *Social Networks and the Semantic Web Series: Semantic Web and Beyond*, Vol. 5 (2007)
- [20] Pretorius, A.J., Lexon visualization: visualizing binary fact types in ontology bases. In proceedings of the 8th International Conference on Information Visualizations. Vol. 14-16 July Washington,DC,USA.(2004)58-63
- [21] Rajput, Q., Haider, S., Tauheed N.: Information Extraction from Unstructured and Ungrammatical Data Sources for Semantic Annotation. Submitted to International Conference on Ontology and Semantic Engineering, Rome, Italy, April (2009)
- [22] Ramoni, M., and Sebastani, P. (1998), *Parameter Estimation in Bayesian Networks from Incomplete Databases*, *Intelligent Data Analysis*, 2.

- [23] Ramoni, M., and Sebastiani, P. (1995), "Learning Bayesian Networks from Incomplete Databases," Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence.
- [24] Tang, J., Li, J., Lu, H., Liang, B., Huang, X., Wang, K.: *IASA: Learning to Annotate the Semantic Web*. Journal on Data Semantics. Vol. 4. (2005) 110-145
- [25] Tjoa, A., Wagner, R., Andjomshoa, A., Shayeganfar, F.: Semantic Web: Challenges and New Requirements. In Proceedings. Sixteenth International Workshop on Database and Expert Systems Application (DEXA). Copenhagen Denmark (2005) 1160 – 1163
- [26] Vargas-Vera, M., Motta, E., Domingue, J., Lanzoni, M., Stutt, A., Ciravegna, F: MnM: Ontology Driven Semi-Automatic and Automatic Support for Semantic Markup. In Proceedings of The 13th International Conference on Knowledge Engineering and Management. Seguenza Spain (2002)
- [27] Wilson, M., Matthews, B.: The Semantic Web: Prospects And Challenges. In Proceeding of 7th International Baltic Conference on Databases and Information Systems. Vilnius Lithuania (2006)
- [28] Yildiz, B., Miksch, S.: Motivating ontology-driven information extraction. In Prasad, A., Madalli, D., eds.: International Conference on Semantic Web and Digital Libraries. Indian Statistical Institute Platinum Jubilee Conference Series (2007) 45–53
- [29] Yildiz Burcu, Miksch Silvia. ontoX - A Method for Ontology-Driven Information Extraction. In: Computational Science and Its Applications (ICCSA 2007), LNCS 4707, Springer-Verlag, 2007, S. 660 - 673.
- [30] Mittal, A. and Kassim, A., *Bayesian Network Technologies: Applications and Graphical Models*, IGI Publishing (2007)
- [31] Pourret, O., Naim, Patrick, and Marcot, B., *Bayesian Networks: A Practical Guide to Applications*, Wiley (2008)
- [32] Heckerman, D. A Tutorial on Learning in Bayesian Networks, Learning in Graphical Models, MIT Press (1999).