Corvinus University of Budapest

Doctoral School of Business

Informatics

Ph.D. Thesis

Summary

Saira Andleeb Gillani

From Text Mining To Knowledge Mining: An Integrated Framework of Concept Extraction and Categorization for Domain Ontology

Supervisor: dr Andrea Kő, PhD

Budapest, 2015

Department of Information Systems

From Text Mining To Knowledge Mining: An Integrated Framework of Concept Extraction and Categorization for Domain Ontology

Ph.D. Thesis

Summary

@ Saira Andleeb Gillani

Supervisor: Andrea Kő, PhD

Budapest, 2015

Contents

I. BACKGROUND AND OVERVIEW OF THE RESEARCH

Organizations are struggling with the challenges coming from the regulatory, social and economic environment which are complex and changing continuously. They cause increase demand for the management of organizational knowledge, like how to provide employees, the necessary job-specific knowledge in right time and in right format. Employees have to update their knowledge, improve their competencies continuously. Knowledge repositories have key roles from knowledge management aspects, because they contain primarily the organizations' intellectual assets (it is explicit knowledge) while employees have tacit knowledge, which is difficult to extract and codify. Business processes are also important from the management of organizational knowledge aspects, they have explicit and tacit knowledge elements as well. One of the key questions is how to handle this hidden knowledge in order to improve the organizational knowledge especially employees' knowledge by providing the most appropriate learning and/or training materials and how can we ensure that the knowledge in business processes are the same as in knowledge repositories and employees' head.

In my thesis emphasis is given to extract information/knowledge from the processes and enrich this organizational information/knowledge by such a paradigm that can disclose the concealed information and interesting patterns by means of methods which dynamically handle the processes' data. A business process splits into tasks and a task has a number of attributes such as description, responsibility, execution related information (order, triggers, and events) and resources to be used. The overall question to be answered in this thesis: what to do in order to achieve the output (new concepts for ontology population) from the input (process description)? The emphasis is on the description attribute of a task because it contains explicit and tacit knowledge elements in an embedded way. My research concern is to identify and mine the hidden knowledge elements and put them into a context (domain ontology) to make it knowledge. Text mining is a method or approach of the extraction of new, heretofore unknown information from any text. Through text mining, description can be extracted from a task and then this description can be used to create such a meaningful list or set of concepts which can be applied as an input for contextualization.

The overall purpose of this dissertation is to propose a solution to transform embedded knowledge of processes in such a form that will be mapped on an ontology. To use process information for domain ontology population, there is a need to perform

advanced analysis of the description (in text form) associated with each activity in log. As it looks from the brief description, the nature of the process modelling is rather procedural, while many questions raised by the modelling need answers on contextual basis, where the context has a rather declarative nature (ontology). This declarative nature is provided in my research by Studio ontology (Vas, 2007).

Structure of the Thesis

The structure of the dissertation is presented in this section. The first chapter deals with motivation, problem statement, research questions and research methodology. I gave an overview about the premise of my work, about the methodology being applied. My main research questions and statements are discussed. The second chapter is about the theoretical background of my work, so I give a detailed description about text mining, business process management, information and concept extraction and ontology learning. The third chapter is describing the proposed framework, ProMine and explain the detailed functionality of different modules of ProMine. Research questions and statements are proved and detailed in the following chapters. Three case studies are examined and evaluated to answer the research questions that frame this study. Chapter seven describes the conclusions and future work. At the end, a summary and contribution of my research work is also given.

I.1 Objectives and Main Questions of the Research

The context of my research is the maintenance of the organizational knowledge. There is an explicit form of this knowledge in knowledge repositories, ontologies and a tacit one in the head of employees. A special knowledge representation form is the business process, which contains knowledge in embedded way. My text mining based knowledge management solution help to connect these knowledge representation forms to each other. One of the key contributions of my text mining solution is to automate the whole process, namely knowledge extraction from business process and to facilitate domain ontology enrichment If there is any alteration in a process or in the knowledge required by the task we throw to modify only the process on high level, and this new or changed knowledge can be easily presented to the employees.

The aim of the thesis is represented by the identification of methods and techniques used for knowledge extraction from organizational process and domain related documents available in digital form.

This research underscores the importance of concept extraction from business processes and ontology learning. Text mining solutions identifies the relationship between activities and job-specific knowledge, soft skills (usually called competencies). The system of job competencies, and organizational structure and business processes can be paired with each other. Relations between concepts which are identified through association rules constitute the basis for ontology evolution. Thus, the output from this text mining framework is expected to be helpful in the maturation of an ontology which will constitute the foundation of the content structure. So, this knowledge will hold employees to easily take their job-role specific knowledge.

I designed and developed a text mining solution, called ProMine, to answer the research questions, which I detail below (code is attached to the thesis). The novelty of the ProMine solution is based on the analysis of business processes' tasks with the help of text mining techniques to extract knowledge from them and in order to connect these business processes to organizational knowledge base, where the process structure will be used for building up the knowledge structure. In this approach knowledge base is the ontology, which provides the conceptualization of a certain domain. The primary innovation lies in new algorithms for the extraction and enrichment of new domain concepts and integration of the static and dynamic process knowledge.

According to these research challenges, this study posits and answers the following research questions:

Research question 1: How can we use text mining to extract knowledge from processes in order to enhance or populate the existing ontology?

This is the main central question of this research and other remaining questions are sub questions of this question. The main challenge of this question is to present a text mining solution/framework that build bridges between two different approaches; process modelling that is procedural in nature and context/ontology is declarative in nature. To answer this question, a thorough literature survey has been studied to see the usefulness of text mining for our required solution. Every perspective has its procedures and the knowledge behind them. The challenge lies in a systematic and gapless integration of these viewpoints.

Research question 2: what methodology and concept extraction methods are available

to enhance the existing knowledge that captured from the business process? Whether a text mining based solution can be used for ontology learning in the context of machine learning?

This is the associate sub question of the first question and it is dealing with my main research issue; discussion of knowledge extraction methods from business processes. Answering this question starts with clarifying how can we articulate the hidden knowledge from business processes by using different text mining techniques. I reviewed theoretical foundations of related fields, like business process management, process mining, text mining, semantic technology and ontologies.

In my thesis emphasis is given to extract organizational knowledge through text mining solution. I have examined different already proposed extraction tools. It is difficult to use existing information extraction methods to enhance the existing knowledge of business process for domain ontology learning. It becomes more appalling when we also want to use this enhanced information again to enrich these business processes. Therefore, there is an increasing need to propose some method or approach in order to automate extraction of information from business processes and enhance this extracted information in such a way that can contribute to enrich existing domain ontologies. The research question is also dealing with the possibilities of the knowledge extraction automation. A variety of text mining approaches have been devised to address the concept extraction problem. Successful techniques include rule based (Mykowiecka, Marciniak, & Kupść, 2009; Xu et al., 2010) and statistical methods (Bunescu et al., 2005; Wu & Weld, 2007). There are different ontology learning approaches such as Text2Onto tool (Cimiano & Völker, 2005), OntoLT (Buitelaar, Olejnik, & Sintek, 2003), OntoBuilder (Gal, Modica, & Jamil, 2004), DODDLE-OWL (Morita, Fukuta, Izumi, & Yamaguchi, 2006) and OntoGen tool (Fortuna, Grobelnik, & Mladenic, 2007) were developed which semi automatically extract concepts for ontology. However, the quality of extracted concepts is low in the previously mentioned solutions as those concepts do not represent the domain well and mostly approaches used traditional ranking metrics (e.g., TF-IDF) thus, they do not show promising results. There is need to use some latest text mining approaches to extract and rank the new concepts.

Research question 3: A modified semantic similarity measure will improve significantly the efficiency and quality of a domain ontology enhancement/ enrichment.

In this thesis, the main goal is the concept extraction from unstructured data (in form of text) of business processes. In literature, some general frameworks are introduced for concept extractions which are less flexible and adaptable. These ontology engineering frameworks rely on shallow NLP techniques for concept extraction. Their emphasis is not on contextual information extraction. Consequently, they neglect to handle semantic phenomena and resulted concepts from such schemes are overly cosmopolitan. From a user perspective, concepts in a context become more meaningful to take decisions. Some state of the art techniques (Ercan & Cicekli, 2007; Gelfand, Wulfekuler, & Punch, 1998; Hassanpour, O'Connor, & Das, 2013; Kok & Domingos, 2008) proposed for semantic concept extraction. A semantic similarity detection technique can allow additional matches to be found for specific concepts not already present in knowledge bases. It is believed that measures of semantic similarity and relatedness can improve the performance in form of quality of such systems. However, these past semantic-based methods, fall short in resolving the main issue: helping users to identify specific concepts related to any business process. It becomes more difficult when there is a reduced quantity of text like, in my case where business processes have not enough domain related data. Therefore, semantic concept extraction is still an open issue in ontology construction and there is a need to implement NLP and text mining techniques in more detail. I propose a new semantic similarity measure which will help in concept extraction and that will overcome the problems of existing semantic similarity measures.

Research question 4: whether taking top categories from the existing ontology will improve the result of text mining solution to help ontology enrichment process or not?

Manual ontology population and enrichment is a complex and time-consuming task that require professional experience involving a lot of expert discussions and efforts. In this thesis, my concern is to propose a semi-automatic solution for ontology population and enrichment. Various approaches based on information extraction methods have already been used for ontology population. Ontology learning, enrichment and maintenance is an ongoing and complex process, with several challenges (Shamsfard & Abdollahzadeh Barforoush, 2003; Wong, Liu, & Bennamoun, 2012; Zouaq, Gasevic, & Hatala, 2011). It has a key role in ontology management; it tackles the issues to turn facts and patterns from the content into shareable high-level constructs or ontologies.

An interesting aspect of ontology population, which is not addressed adequately in the literature, is the handling of redundancy. If an ontology is populated with an instance without checking if the real object or event represented by the instance already exists in the ontology, then redundant instances will be inserted. Therefore, consistency maintenance or redundancy elimination are main issues in ontology population. My proposed solution will handle all these issues by using lexical resources.

II. METHOD OF THE RESEARCH

In reviewing my thesis research methodology in the process of writing this thesis and carrying out research, I had to follow the pattern of traditional research methodologies which is the requirement of the PhD School as well as explored models to perform the research in the domain of computing. A complete research process in solvable tasks has been defined in this thesis and these tasks are time depended and measurable. In order to achieve these solvable tasks, I had to define the problem statement and research questions instead of devising hypothesis.

The Business Informatics Ph.D. School of Budapest Corvinus University belongs to the doctoral schools of social sciences in the university and has been classified to the IT discipline as well, therefore applying research methods in a kind of 'hybrid' way can hopefully be considered to be accepted.

II.1 Proposed Research Methodology

To solve afore mentioned research problem, I have developed such a research methodology that consists of existing processes and findings from design research. This defined process combines qualitative and quantitative research methods. The main design research phases applied in this thesis are as follows;

- **Problem Awareness:** This involves reviewing the literature to analyze the existing techniques of knowledge extraction and ontology enrichment. This phase also confirms the lack of a general framework that can automatically extract knowledge from organizational processes and then after enriching this information use it for ontology enrichment.
- **Extensive Literature Review:** in this phase a thorough literature study took up and identification of unresolved publications of relevant subjects. The purpose of this phase is a detail study about knowledge extraction methods ontology enrichment and concept categorization is a part of this phase and patterns and ontology learning and also level out some issues of these arenas. Nevertheless, the research of this dissertation was a part of an ongoing project (PROKEX project, EUREKA_HU_12-1-2012-0039), so it was intended to keep exploring different techniques/domains to cope with changing requirements/conditions.

- **Developing Research Questions:** After a thorough literature survey, the next step is to develop research questions from the general purpose statement. The focus is narrow down to specific questions. These questions to be answered in later study. This is very important phase because these questions act like driving force behind the research from beginning to end. Four research questions have been developed in this study in which first question RQ1 is the central question and other three are associated sub-questions.
- **Conceptual Framework:** conceptual framework helps to clarify and map out the key research issues in the research area. A tentative idea is produced in this phase for further research. This conceptual mapping is in a pictorial form. Here, it is decided that how to select suitable preprocessing techniques on extracted data from processes and how to apply knowledge extraction techniques. This idea should suggest that how can we use text mining techniques for information extraction, how we can use different machine learning techniques for concept enrichment by using different out resources and how to filter out irrelevant concepts by using different statistical or semantic measures. A detailed conceptual framework has been projected after brainstorming and reviewing sessions.
- **Development:** The development of the solution is achieved by building the design artefact. Here, intend was to develop a working prototype using JAVA that worked as a module within the said project. By immersing in the build activity of this prototype the understanding of the problem becomes more clarify and new suggestions come to mind that helps to improve the next build and evaluate cycle. I have implemented this proposed prototype ProMine with basic modules of the conceptual framework that is for knowledge extraction by using JAVA.
- **Evaluation:** in this phase an assessment method is to develop to assess the quality and effectiveness of the designed artefact (March & Smith, 1995). The purpose of evaluation phase is to consider that what actually occurred. Whether the development met the expected results or not and also check the links between the program as it was delivered and the outcome of the program (Balbach, 1999). The proposed framework, ProMine, is evaluated for coverage of the domain and for accuracy. Both qualitative and quantitative evaluation methods have been established to quantify the performance of the proposed framework. For qualitative evaluation method I have selected three cases from three different domains: 1) Food Chain Safety, 2) Insurance, 3) IT audit. The

understanding of selecting these domains is the Byzantine complexity of interrelated tasks and the problems occurring during their everyday performance. My proposed framework performance is evaluated through controlled experiments with the help of domain experts and applied these results on existing manually engineered ontology (STUDIO).

- **Conclusions:** This is the final phase of the Design Research cycle that covers the overall contribution made by the research. In literature, this phase is given different names like conclusion, results analysis or communication. Conclusions should appropriately supported by evidence. Limitations of my proposed solution and future work are also presented in this research part.

By above mentioned research method, the prototype (ProMine) is prepared. In order to deliver the final version of ProMine, Design Science approach (March & Smith, 1995) have been used. According to this approach, there are two basic activities, build and evaluate; build is the process of constructing and artifact for a specific reason and evaluation is the process of determining the performance of the artifact. In this research, these activities are executed in an iterative incremental Design Research manner consisting of number of iterations as follows:

- Iteration 1 – Core framework development is done including key term extraction technique and concept enrichment module by using WordNet. For concept filtering, statistical measure, information gain is used. Evaluate the technique and tool by using real dataset and evaluating the extracted concepts with the identified evaluation metrics.
- Iteration 2 – For good results of concept extraction, extending the framework and add Wiktionary along with WordNet. For concept filtering, I have devised a semantic measure with the combination of statistical measure. This has improved the results in better form. A link is also created between extracted concepts and business processes.
- Iteration 3 – In third iteration, I have improved the concept extraction technique by modifying compound word selection method and the main development of this iteration was the concept categorization module that is be prepared by using seed ontology. This categorization module provides help in ontology enrichment.
- Iteration 4 – Validate the framework by applying and evaluating the extraction method across other domains. The generality of the framework and ProMine tool is demonstrated through comparing evaluation measures for three different domains.

Through this iterative procedure of the text mining prototype, three real case scenarios have been run to illustrate the effectiveness and provide a live proof of the proposed method (ProMine) and as the means by efficiencies and improvements are identified. Determining whether progress is made by the extraction method and tool is evaluated by applying the appropriate metrics from the knowledge base to measure the accuracy and coverage of the learned domain ontology model.

III. PROMINE: THE PROPOSED FRAMEWORK

I have developed a prototype workbench that performs three basic tasks; one is knowledge element extraction from the domain document corpus and other sources, and then concept filtering to find most relevant terms of a domain from the extracted knowledge elements. The third task is semantic concept categorization with these extracted knowledge elements that will help the enrichment and population of domain ontology. Initial work of ProMine concept extraction is presented in (Gillani & Kő, 2014). This prototype shows that proposed framework's efficacy as a workbench for testing and evaluating semantic concept extraction, filtering and categorization.

The workflow of the proposed ontology framework proceeds through the following phases:

III.1 Data Extraction

ProMine's input file is actually the output file of an organizational process by using a process model. This input file is in the form of XML. At the first step of this framework (data extraction phase), the pertinent information from this input file is extracted automatically by ProMine. After extracting specific text from the input files, this text is saved into text files according to all tasks.

III.2 Preprocessing of Data

After text extraction, the most crucial part, cleaning of extracted text starts. Preprocessing portrays any sort of transformation performed on unstructured text to set it up in such format that it will be easily and efficiently processed. This preprocessing module ensures that data are prepared for subsequent activities, which are discussed later in the description. Text preprocessing is an integral part of natural language processing (NLP) system. Text preprocessing include in general different NLP and text mining techniques such as tokenization, stop word removal, part-of-speech (POS) tagging, stemming or lemmatization and frequency count. By applying these techniques, the input text is transformed into term vector and the weight of each term is based on the frequency of the term in an input file.

III.3 Concept Enrichment

At the end of preliminary phase, a set of unique key words is created against each

organizational task. This phase can be divided into two steps; first step extracted synonyms from different lexical resources and in second step compound words are made with the help of domain corpus.

III.4 Concept Filtering based on Semantic Similarity Measure

Though, till last phase unrelated terms (conceptually, not related to a specific domain) from a set of synonyms terms (from WordNet & Wiktionary) of a given key term has been removed. However, the resultant word list consists of lexical terms which are hundreds in number. This high dimensionality of the feature space is the major particularity of text domain. These unique concepts or potential concepts are considered as feature space, these lists of concepts can be considered as high dimensional and sparse vectors. In our proposed framework, at this stage, I am reducing feature space by selecting more informative concepts from this concept list by using a concept filtering method. Conventionally, in most ontology learning tools, statistical measures such as TF-IDF, RTF, entropy or probability methods are used for filtering process (Cimiano & Völker, 2005). To identify important lexical terms, ProMine used an innovative approach that is the combination of statistical and semantical measures. We have proposed a new hybrid semantic similarity measure to identify relevant ontological structures for a given organizational process. This module consists of two phases; in first phase for each candidate concept its information gain (IG) is calculated by using domain corpus and in second phase to find more semantically representative candidate concepts we proposed our hybrid semantic similarity measure that uses different information sources such as lexical semantic network (WordNet) and domain corpus.

III.5 Semantic Concept Categorization

ProMine architecture performs basically two main tasks; one is concept extraction and the second is semantic concept categorization. At the end of third phase, we have a refined list of domain specific concepts against each key term that was selected in the second phase. These extracted concepts are semantically similar to the key term. Now, we want to categorize these concepts in such a manner that ontology to be enriched. For this purpose, it is necessary to find out concept relationships between these words and existing (seed) domain ontology. We proposed a novel semantic concept categorization

method to enrich an existing ontology. This method will classify new domain-specific concepts according to the existing taxonomy of the seed ontology. For concept categorization, this method will use the knowledge of existing concept categories (taxonomy of classes) of the ontology with the help of external knowledge resources such as Wiktionary. The proposed approach tries to find a semantic similarity between extracted concepts using some fragment of the ontology that describes a certain category.

IV. RESULTS OF THE RESEARCH, CONTRIBUTION OF THE THESIS

In this dissertation, I presented a new paradigm of text mining with respect to business processes/organizations. I have developed ProMine prototype workbench that performs these text mining tasks. Through this text mining solution, I had to identify and mine the hidden knowledge elements from business processes and put them into a context (domain ontology) to make it knowledge.

From three case studies, one by one, I strived to answer all research questions. Results of first case study showed that ProMine concept extraction approach can support domain experts and ontology engineers in building domain specific ontologies efficiently. However, to reduce rate of concepts that should be in final concept list, some other methods besides information gain should be also considered. As a weakness of the IG criterion is that it is biased in favor of features with more values even when they are not more informative (Novaković, Štrbac, & Bulatović, 2011) and as I mentioned earlier this is pure statistical measure so we have to consider some semantic measures as well for better results. Therefore, I proposed a new hybrid similarity measure (Cloud Kernel Estimator) for concept filtering process.

The second case study is about the proposed similarity measure. Besides concept extraction ProMine also addresses ranking and filtering relevant terms by using a new hybrid similarity measure. The novelty of this extracting tool is that 1) it extracts concepts from a very little knowledge that are embedded in organizational processes and with the help of outsources (WordNet, Wiktionary) enrich this knowledge and extracts a huge number of new concepts automatically without human interaction; 2) its filtering approach uses deep syntactic and semantic analysis to filter important concepts. The proposed new hybrid similarity measure can be used for other applications of artificial intelligence, psychology and cognitive science. This case study illustrated that the performance of ProMine was assessed using a human evaluation and the results showed that many new concepts were successfully extracted and later on used for ontology population. In third case study, ProMine ontology enrichment capability is tested in an ontology-based e-learning environment applied for IT auditors' training in CISA preparation courses. An ideal system with high precision and high recall will return many results, with all results labelled correctly. As the number of cycles passed, ProMine system showed good precision in enriching the ontology, a perspective of the

work is to extend the number of cycles. If we see precision and recall both then we will come to know that both are increasing gradually. High scores for both showed that the proposed categorization method is returning accurate results (high precision), as well as returning a majority of all positive results (high recall). The experimental evaluation using precision and recall measures as well human judgments showed that ProMine method categorized new concepts with high precision and recall.

The main contributions of this dissertation are:

- A key contribution of this thesis is to provide a generic text mining solution/framework that build bridges between two different approaches; process modelling that is procedural in nature and context/ontology is declarative in nature.
- The major contribution is concept extraction and enrichment with the help our resources such as WordNet, Wiktionary and domain corpus. The state-of-the-art text mining and NLP techniques are used in information extraction solution.
- Developing of a generic method of discovering useful knowledge in terms of single key term and compound terms to some key issues discussed in the textual databases.
- For concept filtration, I proposed a new method which is a combination of statistical and semantic measures. Through this process more semantic and contextual concepts of any sphere can be elicited from a given text data. This proposed method is a hybrid similarity measure.
- Another important contribution is to design a concept categorization method for ontology population.
- The proposal of novel integration of text mining techniques to capture knowledge elements from organizational processes and disseminate these elements in terms of new concepts in domain ontology.

The ProMine solution has more potential for the future development. Comparison analyses of different concept extracting and filtering methods have not been considered during the current research work, but this could also help to improve the methodological development of the proposed framework, as a part of future research work. The current research has focused on concept extraction methods whereas the relation (between concept) extraction could be used to improve the ontology enrichment as a part of future research work.

V. REFERENCES

Balbach, E. D. (1999). Using case studies to do program evaluation. *Sacramento, CA: California Department of Health Services*.

Buitelaar, P., Olejnik, D., & Sintek, M. (2003). *OntoLT: A protege plug-in for ontology extractionfrom text.* Paper presented at the Proceedings of the International Semantic Web Conference (ISWC).

Bunescu, R., Ge, R., Kate, R. J., Marcotte, E. M., Mooney, R. J., Ramani, A. K., & Wong, Y. W. (2005). Comparative experiments on learning information extractors for proteins and their interactions. *Artificial intelligence in medicine, 33*(2), 139-155.

Cimiano, P., & Völker, J. (2005). Text2Onto *Natural language processing and information systems* (pp. 227-238): Springer.

Ercan, G., & Cicekli, I. (2007). Using lexical chains for keyword extraction. *Information Processing & Management, 43*(6), 1705-1714.

Fortuna, B., Grobelnik, M., & Mladenic, D. (2007). *OntoGen: semi-automatic ontology editor*: Springer.

Gal, A., Modica, G., & Jamil, H. (2004). *Ontobuilder: Fully automatic extraction and consolidation of ontologies from web sources.* Paper presented at the Data Engineering, 2004. Proceedings. 20th International Conference on.

Gelfand, B., Wulfekuler, M., & Punch, W. (1998). *Automated concept extraction from plain text.* Paper presented at the AAAI 1998 Workshop on Text Categorization.

Gillani, S. A., & Kő, A. (2014). Process-based knowledge extraction in a public authority: A text mining approach *Electronic Government and the Information Systems Perspective* (pp. 91-103): Springer.

Hassanpour, S., O'Connor, M. J., & Das, A. K. (2013). A semantic-based method for extracting concept definitions from scientific publications: evaluation in the autism phenotype domain. *J. Biomedical Semantics, 4*, 14.

Kok, S., & Domingos, P. (2008). Extracting semantic networks from text via relational clustering *Machine Learning and Knowledge Discovery in Databases* (pp. 624-639): Springer.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision support systems, 15*(4), 251-266.

Morita, T., Fukuta, N., Izumi, N., & Yamaguchi, T. (2006). DODDLE-OWL: a domain ontology construction tool with OWL *The Semantic Web–ASWC 2006* (pp. 537-551): Springer.

Mykowiecka, A., Marciniak, M., & Kupść, A. (2009). Rule-based information extraction from patients' clinical data. *Journal of biomedical informatics, 42*(5), 923-936.

Novaković, J., Štrbac, P., & Bulatović, D. (2011). Toward optimal feature selection using ranking methods and classification algorithms. *Yugoslav Journal of Operations Research ISSN: 0354-0243 EISSN: 2334-6043, 21*(1).

Shamsfard, M., & Abdollahzadeh Barforoush, A. (2003). The state of the art in ontology learning: a framework for comparison. *The Knowledge Engineering Review, 18*(04), 293-316.

Wong, W., Liu, W., & Bennamoun, M. (2012). Ontology learning from text: A look back and into the future. *ACM Computing Surveys (CSUR), 44*(4), 20.

Wu, F., & Weld, D. S. (2007). *Autonomously semantifying wikipedia.* Paper presented at the Proceedings of the sixteenth ACM conference on Conference on information and knowledge management.

Xu, H., Stenner, S. P., Doan, S., Johnson, K. B., Waitman, L. R., & Denny, J. C. (2010). MedEx: a medication information extraction system for clinical narratives. *Journal of the American Medical Informatics Association, 17*(1), 19-24.

Zouaq, A., Gasevic, D., & Hatala, M. (2011). Towards open ontology learning and filtering. *Information Systems, 36*(7), 1064-1081.

VI. REFERENCES OF THE AUTHOR

2015 **Saira Gillani**, Andrea Kő

Incremental Ontology Population and Enrichment through Semantic-based Text Mining: An Application for IT Audit Domain.

International Journal on Semantic Web and Information Systems (IJSWIS). (Under Second Review).

ISSN: 1552-6283, DOI:10.4018/IJSWIS

2014 **Saira Gillani**, Andrea Kő

Process-based Knowledge Extraction in a Public Authority: A Text Mining Approach

2014 3rd International Conference (EGOVIS 2014) on Electronic Government and the Information Systems Perspective, Springer International Publishing.

ISBN: 978-3-319-10178-1, pp. 91-103, DOI: 10.1007/978-3-319-10178-1

2015 Noureen Zafar, Saif Ur Rehman, **Saira Gillani,** S.Asghar

Segmentation of Crops and Weeds using Supervised Learning Technique.

Improving Knowledge Discovery through the Integration of Data Mining Techniques., Book Chapter, Advances in Data Mining and Database Management (ADMDM) Book Series, IGI Global.

ISBN: 9781466685130, DOI:10.4018/978-1-4666-8513-0

2015 Mohsin Iqbal, Saif Ur Rehman, **Saira Gillani,** S. Asghar

An Empirical Evaluation of Feature Selection Methods.

Improving Knowledge Discovery through the Integration of Data Mining Techniques., Book Chapter, Advances in Data Mining and Database Management (ADMDM) Book Series, IGI Global.

ISBN: 9781466685130, DOI:10.4018/978-1-4666-8513-0

2013 Muhammad Naeem, **Saira Gillani** and Sheneela Naz,

MfWMA: A Novel Web Mining Architecture for Expert Discovery

International Journal of Advanced Science and Technology (IJAST).

ISSN: 2005-4238, Vol. 52, March, 2013.

2013 **Saira Gillani**, Peer Azmat Shah, Amir Qayyum, Halabi Hasbullah

MAC Layer Challenges and Proposed Protocols for Vehicular Ad-hoc Networks

In Vehicular Ad-hoc Networks for Smart Cities, Springer International Publishing.

ISBN: 978-981-287-158-9, 10.1007/978-981-287-158-9_1, pp. 3-13.

2013 **Saira Gillani,** Farrukh Shahzad, Amir Qayyum and Rashid Mehmood

A Survey on Security in Vehicular Ad hoc Networks

In LNCS series (Springer) Nets4Cars-Nets4Trains

ISBN: 978-3-642-37974-1, DOI:10.1007/978-3-642-37974-1_5

2013 **Saira Gillani,** Muhammad Naeem, Raja Habibullah and Amir Qayyum

Semantic Schema Matching Using DBpedia

International Journal of Intelligent Systems and Applications (IJISA)

ISSN: 2074-9058, Vol. 5, No. 4

2012 **Saira Gillani**, Shahnela Naz, Muhammad Naeem, Tanvir Afzal, Amir Qayyum

ErraGMap: Visualization Tool

In 8th International Conference on Digital Content Technology and its Applications IDCTA2012, Jeju Island, Korea.

2008 **Saira Gillani,** Imran Khan**,** Shahid Qureshi, Amir Qayyum

Vehicular Ad hoc Network (VANET), Enabling Secure and Efficient Transportation System

In Technical Journal, University of Engineering and Technology, Taxila, Pakistan. Dec.

ISSN: 1813–1786, Volume 13.