DISCOVERING RELATIONSHIP ASSOCIATIONS FROM THE LITERATURE RELATED TO RESEARCH PROJECTS IN SUSTAINABILITY SCIENCE USING ONTOLOGY AND INFERENCE

Weisen Guo and Steven B. Kraines
Division of Project Coordination, the University of Tokyo, Kashiwa-No-Ha, Kashiwa-Shi, Japan

Keywords: Sustainability, Relationship Association, Ontology, Inference, Literature, Research Project.

Abstract: Research projects addressing issues related to sustainability often need knowledge from research papers from a wide range of disciplines. A method is developed and assessed for using ontology-based inference to automatically discover knowledge in semantic statements of research papers related to specific research projects in sustainability science. The semantic statements have been constructed using a semi-automatic authoring process to represent the knowledge content of the research papers. The discovered knowledge is expressed in the form of relationship associations that are extracted from semantic statements, where relationship associations are transitive associations between two binary semantically typed relationships that share a connecting entity and that co-occur frequently in the set of semantic statements. An algorithm is presented here for finding interesting relationship associations that are extracted from research papers and related to a given research project. The method is evaluated on a set of semantic statements containing 104 semantic statements describing research papers and 24 semantic statements describing research projects.

1 INTRODUCTION

The emerging field of sustainability science faces the challenge of aligning global development with the resource and waste processing limits of the Earth in such a way that can be sustained for the foreseeable future. Many research projects in sustainability science are being conducted, and these projects require specialized knowledge from a wide range of different scientific, sociological, and economic domains. On the other hand, the rate at which knowledge resources such as research papers are published in these different domains is increasing exponentially. Therefore, how to discover pre-existing knowledge from different domains that is both interesting and relevant to particular research projects in sustainability science is a critical problem.

This paper presents a method to use ontology-based inference to discover interesting knowledge from the literature that is related to specific research projects in sustainability science. The method is based on "computer understandable" knowledge descriptors created semi-automatically to describe the knowledge content of a research paper or project using terms and relationships provided by an OWL-DL ontology. The knowledge descriptors, which we call semantic statements, are computer-understandable in the sense that a computer can infer new facts and implications from the descriptors using logic and rules (Kraines et al., 2006).

The discovered knowledge is expressed in the form of relationship associations that are extracted from the descriptors (Guo and Kraines, 2009). Relationship associations are pairs of binary typed and directed relationships, often called semantic triples and occurring in the form "subject – verb -> object", that share a connecting entity and that co-occur frequently in the set of semantic statements.

The method reported here attempts to find the relationship associations extracted from papers in the literature that are relevant to a given semantic statement for a research project. In short, we look for relationship associations occurring in at least two research papers, where the first relationship in the association occurs in the research project semantic statement. These relationship associations are hypothesized to be useful for the researchers involved in the given research project, e.g. to

Guo W. and B. Kraines S..
DISCOVERING RELATIONSHIP ASSOCIATIONS FROM THE LITERATURE RELATED TO RESEARCH PROJECTS IN SUSTAINABILITY SCIENCE USING ONTOLOGY AND INFERENCE.
DOI: 10.5220/0003632603820385
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (KDIR-2011), pages 382-385
ISBN: 978-989-8425-79-9

identify important related topics that they might have overlooked. To test the effectiveness of this knowledge discovery process, we apply it to a set of semantic statements describing research papers and research projects related to sustainability science.

The original contribution of this paper is a method for discovering knowledge in the form of potentially important new relationships for an entity that is the focus of a research project from relationship associations mined from a set of semantic statements representing scientific papers.

This paper is organized as follows. In Section 2, we describe our method for discovering related relationship associations from a set of semantic statements. In Section 3, we present the results of applying the method to a set of semantic statements of research papers and research projects created in other work. In Section 4, we conclude this paper with a summary.

2 MATERIALS AND METHODS

Our goal is to discover knowledge in specific research papers that could be useful to the researchers involved in particular research projects in sustainability science.

The first step in our proposed knowledge discovery method is to mine the set of semantic statements, which have been created to describe the knowledge content of individual research papers, for interesting relationship associations. One relationship association contains an ordered pair of semantic triples that share one common entity, where a semantic triple is comprised of a subject entity plus a directed and typed relationship plus an object entity. A relationship association of the form (e1 –r1-> e2, e1 –r2-> e3) can be interpreted as follows: if a particular entity e1 has a relationship r1 with another entity e2, then it is likely that e1 has another relationship r2 with a third entity e3.

For this mining step, we use an algorithm developed in previous work (Guo and Kraines, 2009; Guo and Kraines, 2010; Guo and Kraines, 2011). Briefly, the algorithm works as follows. The input of the algorithm is a set of semantic statements, in this case all of the semantic statements of research papers. The output of the algorithm is a set of interesting relationship associations in the form of linked directed pairs of semantic triples. First, we extract all of the semantic triples from the set of semantic statements and convert the triples to triple queries. Second, we obtain the support for each triple query in the entire set of semantic statements, and we discard queries that do not meet a support threshold. Third, we generate relationship associations, which express associations between all pairs of the triple queries that remain. Fourth, we remove relationship associations that are semantically equivalent. Fifth, we obtain the support for each of the remaining relationship associations in the entire set of semantic statements using logic and rule based inference.

Next, we want to discover which of the relationship associations extracted using the mining process described above are related to each semantic statement describing a research project. Because we want relationship associations that are not specific to a single paper, our first requirement is that the relationship association must have a support of at least two, i.e. it must occur in at least two of the semantic statements for the research papers. Due to the relatively small number of semantic statements that we have here (104 statements describing research papers), we do not use the relevance criteria from our previous mining algorithm (Guo and Kraines, 2011). However, if the set of semantic statements is larger, we can apply those criteria as well to reduce the size of related and potentially interesting relationship associations discovered.

We then look for relationship associations where the first relationship occurs in the semantic statement for a research project but where the entire relationship association does not occur. If we find such a relationship association for a particular research project, then we can suggest that the researchers involved in the research project consider the second relationship as knowledge that they might want to include in their description of the project. For example, in one of the projects we are studying, a researcher is studying an energy system that has a fuel cell as a component. We have found the following relationship association from the research papers: "if a **fuel cell** *is part of* an **energy system**, then it is likely that the **energy system** *has quantity* **efficiency**" (hereafter the text in bold font indicates class and the text in italic font indicates property). If the project description did not mention the efficiency of the energy system, we can recommend that the project researcher consider adding that to the project description.

3 RESULTS

We evaluate the method using 24 semantic statements describing research projects that have been funded by the AGS Promotion Office at the

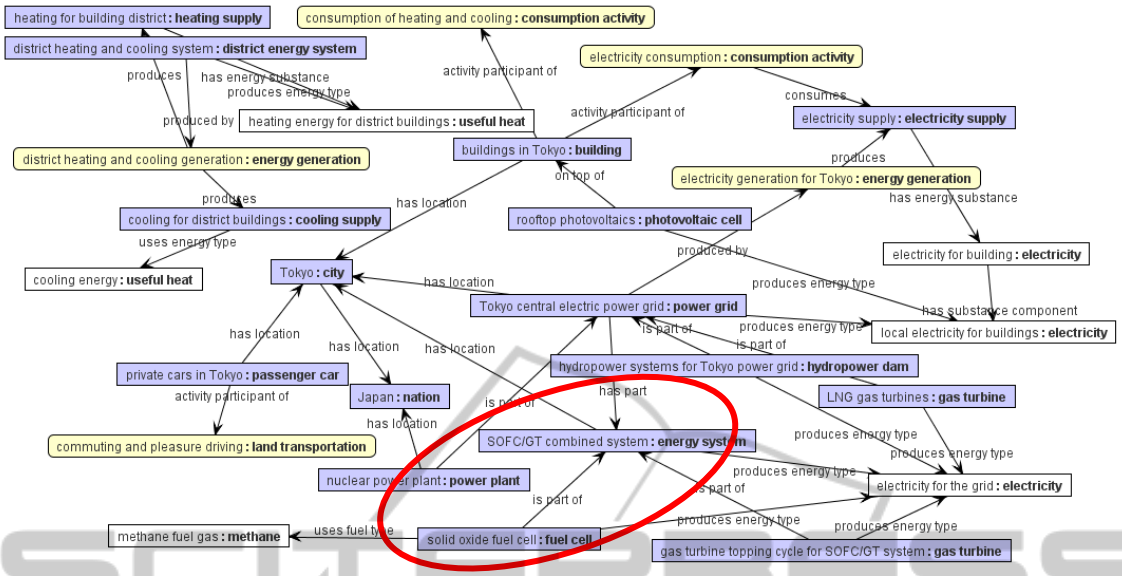

Figure 1: Graph view of the semantic statement for the project entitled "An integrated evaluation system for countermeasure technologies to realize sustainable societies in cities". Instances of ontology classes describing entities from the text abstract are shown with boxes colored according to the major upper class: physical objects are blue, activities are yellow, events are gray, abstract objects are white. The free text name of the instance is followed by a colon and the name of the class. Properties specifying the relationship between the instances are shown as directed arrows labeled with the name of the property.

University of Tokyo between 2000 and 2005 and 104 semantic statements that have been created to describe research papers related to energy and sustainability. All 128 of the semantic statements were created through the EKOSS system (Kraines et al., 2006) using the SCINTENG ontology (Kraines and Guo, 2011) as the knowledge representation language.

From the set of 104 semantic statements describing research papers, we generated 47,419 semantically unique relationship associations. However, only 132 relationship associations occur in the set of semantic statements at least twice. Unlike concept associations, relationship associations are directed. Therefore, we can also consider the reverse of each relationship association, which gives us 264 relationship associations to use in the knowledge discovery step. Using the algorithm presented in the previous section, we discovered that, on average, 10 of the 264 relationship associations were related to each of the 24 research project semantic statements.

One example for the research project entitled "An integrated evaluation system for countermeasure technologies to realize sustainable societies in cities". The graph view of the semantic statement created based on the project abstract is shown in Figure 1. The relationship association that was discovered to be related to this semantic statement is:

"If an **energy system** *has part* **fuel cell**, then it is likely that the **energy system** *has quantity* **efficiency.**"

This relationship association occurs in the semantic statements describing the research paper entitled "Cycle analysis of micro gas turbine-solid oxide fuel cell hybrid system" (Uechi et al., 2002), and the research paper entitled "Cycle analysis of micro gas turbine-molten carbonate fuel cell hybrid system" (Kimijima and Kasagi, 2005).

Based on this relationship association, we can suggest that the project researchers might consider studying the efficiency of the SOFC/GT combined system, which has a solid oxide fuel cell as a part, as shown by the triple circled in red in Figure 1.

Another example is for the research project entitled "development of a strategic traffic model for the Tokyo urban area and application to the analysis of control of environmental loading" (We ignore the graph view of the semantic created based on the project abstract for the space limitation). The relationship association that was discovered to be related to this semantic statement is:

"If a **gaseous phase fluid object** *has material* **co2 pollutant**, then it is likely that an **emission activity** *emits* that **gaseous phase fluid object**."

This relationship association occurs in the semantic statements describing the research paper entitled "Simulation of tradable CO_2 emission permits with the New Earth 21 Model" (Yamaji et al., 1998), and the research paper entitled "Classifying CO_2 emissions from the viewpoint of LCA by reflecting the influence of regional activities" (Yoshikuni et al., 1998).

Based on this relationship association, we suggest that the researchers consider that the vehicle air pollutant emissions result from emission activities that are sub activities of the commuting activities.

4 CONCLUSIONS

Sustainability science is fundamentally multi-disciplinary, and research projects in sustainability science often need knowledge from many different research fields. How to effectively utilize the knowledge existing in the scientific literature to help researchers resolve issues during their research projects is a critical problem. We have described and tested a method to discover knowledge related to specific research projects in the form of relationship associations, which are extracted from computer-understandable descriptors called semantic statements. We evaluated our method using 104 semantic statements describing research papers and 24 semantic statements describing research projects. Several relationship associations that we found to be related to specific research projects appear to suggest aspects of key entities in the research projects that might be of interest to the researchers based on knowledge expressed in the research papers.

ACKNOWLEDGEMENTS

This work was funded by the Alliance for Global Sustainability Promotion Office at University of Tokyo.

REFERENCES

Guo, W., Kraines, S. B., 2009. Discovering Relationship Associations in Life Sciences Using Ontology and Inference, In: *Proceedings of the 1st International Conference on Knowledge Discovery and Information Retrieval 2009, KDIR 2009*, pp:10-17, DOI: 10.5220/0002285300100017, SciTePress.

Guo, W., Kraines, S. B., 2010, Mining Relationship Associations from Knowledge about Failures using Ontology and Inference, In: *Proceedings of the 10th Industrial Conference on Data Mining ICDM 2010*, Berlin, Germany, July 12-14. In: P. Perner (Ed.): *ICDM 2010, LNAI 6171*, pp. 617-631, DOI: 10.1007/978-3-642-14400-4_48, Springer-Verlag Berlin Heidelberg 2010.

Guo, W., Kraines, S. B., 2011. Extracting Relationship Associations from Semantic Graphs in Life Sciences, In: A. Fred et al. (Eds.): *IC3K 2010, CCIS 128*, pp. 53-67, 2011, Springer-Verlag Berlin Heidelberg 2011.

Kimijima, S., Kasagi, N., 2005. Cycle analysis of micro gas turbine-molten carbonate fuel cell hybrid system, *Nippon Kikai Gakkai Ronbunshu, Japan Society of Mechanical Engineers International Journal Series B*, 48(1): 65-74, DOI: 10.1299/jsmeb.48.65.

Kraines, S., Guo, W., Kemper, B., Nakamura, Y., 2006. EKOSS: A Knowledge-User Centered Approach to Knowledge Sharing, Discovery, and Integration on the Semantic Web. In: Cruz, I. et al. (eds) *The Semantic Web – ISWC 2006, LNCS, vol. 4273*, pp 833-846. DOI 10.1007/11926078_60.

Kraines, S. B., Guo, W., 2010. Supporting Reuse of Knowledge of Failures through Ontology-based Semantic Search, In: *Proceedings of the 2nd International Conference on Knowledge Management and Information Sharing, KMIS 2010*, 25-28 October, 2010, Valencia, Spain, pp: 164-169, DOI: 10.5220/0003068001640169, SciTePress.

Kraines, S. B. and Guo, W., 2011. A system for ontology-based sharing of expert knowledge in sustainability science. *Data Science Journal, Vol. 9*: 107-123, 29 January 2011, DOI: 10.2481/dsj.Kraines.

Uechi, H., Kimijima, S., Kasagi, N., 2002. Cycle analysis of micro gas turbine-solid oxide fuel cell hybrid system, *Nippon Kikai Gakkai Ronbunshu, B Hen/Transactions of the Japan Society of Mechanical Engineers, Part B*, 68(666): 626-635.

Yamaji, K., Sato, K., Fujii, Y., Akimoto, K., 1998. Simulation of tradable CO2 emission permits with the New Earth 21 Model, *International Journal of Global Energy Issues*, 11

Yoshikuni, Y., Hisashi, I., Ryuji, M., 1998. Classifying CO2 emissions from the viewpoint of LCA by reflecting the influence of regional activities, *Nihon Enerugi Gakkaishi/Journal of the Japan Institute of Energy*, 77