Data Science Journal, Volume 9, 29 January 2011

A SYSTEM FOR ONTOLOGY-BASED SHARING OF EXPERT KNOWLEDGE IN SUSTAINABILITY SCIENCE

Steven Kraines[1*], *Weisen Guo*[2]

[*1]*Science Integration Programme – Human, Division of Project Coordination, Univ. of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa-shi, Chiba Prefecture, 277-8568, Japan*
Email: sk@scint.dpc.u-tokyo.ac.jp
[2]*Science Integration Programme – Human, Division of Project Coordination, Univ. of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa-shi, Chiba Prefecture, 277-8568, Japan*
Inst. of Systems Eng. of Dalian University of Technology, Dalian 116024, China
Email: gws@scint.dpc.u-tokyo.ac.jp

ABSTRACT

Work towards creation of a knowledge sharing system for sustainability science through the application of semantic data modeling is described. An ontology grounded in description logics was developed based on the ISO 15926 data model to describe three types of sustainability science conceptualizations: situational knowledge, analytic methods, and scenario frameworks. Semantic statements were then created using this ontology to describe expert knowledge expressed in research proposals and papers related to sustainability science and in scenarios for achieving sustainable societies. Semantic matching based on logic and rule-based inference was used to quantify the conceptual overlap of semantic statements. This shows the semantic similarity of topics studied by different researchers in sustainability science, similarities that might be unknown to the researchers themselves.

Keywords: Expert knowledge, Knowledge sharing, Semantic web, Semantic search, Knowledge description, Sustainability science, Ontology

1 INTRODUCTION

We have developed a system based on principles of semantic data modeling for realizing a knowledge sharing platform mediated by a computer searching and matching engine that uses logical and rule-based inference to calculate the semantic similarity between computer-understandable descriptions of the knowledge resources, such as academic papers and databases, which are shared on the platform. The system, called EKOSS for Expert Knowledge Ontology-based Semantic Search, provides authoring tools that empower the authors of the resources to create descriptions using formalized languages that can be interpreted semantically, and thereby “understood”, by computers. Then, using those computer-understandable descriptions, which we call semantic statements, EKOSS provides computer-enabled knowledge sharing services, such as semantic searching and matching of specific relationships between target entities that are not explicitly stated but that can be inferred from the formal structure of the language used to author the semantic statements.

Here, we present application of the EKOSS system to the domain of sustainability science. Drawing on a number of knowledge classifications such as the ISO 15926 data model standard, a formalized knowledge representation language in the form of an ontology grounded in a description logic, called SCINTENG, has been created to describe concepts related to sustainability science in three knowledge domains: situational knowledge, analytic methods, and scenario frameworks. We are using the SCINTENG ontology to create semantic statements describing expert knowledge in sustainability science expressed by research project proposals, research papers, and scenario descriptions. Commercially available reasoning software can be applied to pairs of semantic statements to compute the extent of conceptual overlap in a semantic way, e.g. by accounting for the logical characteristics of specific relationships predicating the way in which subject entities are modified by particular object entities (Guo & Kraines, 2008b). The quantified degree of conceptual overlap can then be used in a number of value-adding tasks, such as evaluating the semantic similarity of the research topics of different researchers.

The paper is structured as follows. In the next section, we review the special characteristics of sustainability science as a transdisciplinary study, and we argue for the effectiveness of an approach for matching expert knowledge resources based on descriptions that are created in a computer-understandable form by the authors of the resources themselves. In section 3, we describe a platform that we have developed as an implementation of this approach, a platform for sharing expert knowledge based on computer-understandable semantic statements created by the human authors of the resources that embody the shared knowledge. In section 4, we report the initial results of this effort. Those results include 1) the development of an ontology as a computer-interpretable knowledge representation language for the domain of sustainability science, 2) the creation of an initial set of semantic statements for different types of knowledge resources in sustainability science, and 3) examples of semantic matches that can be inferred between the semantic statements using computer reasoning software. In section 5, we discuss some of the implications of the envisaged knowledge sharing platform, and we list some future research topics that we are exploring in order to realize the envisaged platform for sustainability science.

The major contributions of this paper are thus a proposal for a formal knowledge representation language in the form of a logic-based ontology that can be used to describe research and scenarios in the domain of sustainability science, a corpus of semantic statements created using this ontology, and a demonstration of how logic and rule-based inference can be applied to accurately match the semantic statements using examples of real scenarios and research papers.

2 BACKGROUND AND RELATED RESEARCH

Transdisciplinary studies are critical for addressing issues of sustainability science (Kostoff, 2005; Kostoff, 2002; Takeuchi & Komiyama, 2006; Clark, 2007). For example, over the past decade, the Alliance for Global Sustainability (AGS) Promotion Office at the University of Tokyo has funded projects covering a large number of academic fields, from chemical engineering to economics to political science to cultural studies (AGS Promotion Office at the University of Tokyo, 2009). However, although the breadth of this range of disciplines is one of the major strengths of multidisciplinary funding organizations such as the AGS, it is also a potentially serious weakness. Without effective technologies for integrating the expert knowledge emerging from these widely differing fields of study, there is a danger that the knowledge deliverables will remain disconnected (Neumann & Prusak, 2007). As the amount of knowledge resources increases, it will become increasingly difficult for anyone to get an overview of the range of knowledge that has been created and understand the potential interrelationships and structure of that knowledge, which could be critical in achieving the goal of sustainability (Takeuchi & Komiyama, 2006).

In fact, integration of knowledge from a wide range of both academic and non-academic domains is often a determining factor in developing solutions for problems of global sustainability (Takeuchi & Komiyama, 2006; Clark, 2007). Consider the problem of conducting a comprehensive environmental impact analysis of the potential for some new technology, such as a fuel-cell-based combined heat and power system, to be actually adopted into a society and to effectively contribute to mitigation of some environmental problem, such as global warming. Many forms of expert knowledge in addition to the technological knowledge of the fuel cell system itself may be necessary to evaluate the feasibility of the large-scale application of the technology, for example, knowledge related to the lifecycles of rare materials required for the construction of the system and knowledge on the effects of large-scale introduction of the system on existing systems for providing the same kinds of services (Kraines et al., 2005; Kraines et al., 2006b; Fukushima et al., 2004).

Information and communication technologies (ICTs), particularly those emerging in the context of data modeling and the Semantic Web, could play an important role in facilitating the integration of large quantities of expert knowledge that cannot be easily covered by hand (Garfield, 2001; Seringhaus & Gerstein, 2008; Uren et al., 2006). Cahlik (2000) used scientometric techniques to generate and compare knowledge maps of different domains of knowledge. Kajikawa et al. (2007) showed how citation data can be used to construct an overview of the domain of sustainability science, and Borner et al. (2003) reviewed a range of techniques that have been used to visualize the results of such large-scale analyses of knowledge repositories. Finally, several large scale systems are being developing to create social networks of researchers that are based on some semantics, such as VIVO, Collexis, and Arnetminer (Gewin, 2009; Gunter, 2009; Tang et al., 2007). These systems use some combination of bibliographic and/or thesaurus-type relations to compute the semantic similarity between researchers and/or research papers.

While these approaches to map existing knowledge repositories such as Web of Science are important first steps towards answering the need for knowledge integration, connections that are identified between researchers using these scientometric techniques, such as networks based on coauthorship or citation links, only tell us who already knows whom. Existing approaches to measuring actual semantic similarity are based on comparisons of “bag-of-word” indices, which cannot evaluate the specific relationships between the terms to which the knowledge refers. It has been often observed that most scientific facts occur not as single concepts but as specific relationships between concepts (Weikum et al., 2009). Establishing meaningful but unknown connections between knowledge resources and their creators, e.g., that tell us whom we should but do not already know, requires an analysis that considers the semantics of these relationships between concepts.

Furthermore, existing scientometric and information extraction techniques do not actively involve the actual knowledge creators, a.k.a. the authors of the research papers and developers of simulation models, who are the most likely to know how best to describe their research in the clear and unambiguous form necessary for computer understanding (Abbott, 2004; Mons et al., 2008). Technologies that enable the creators of knowledge resources to provide computer understandable descriptions of their resources themselves, rather than relying on third-party annotators or automated computer “bots”, could make computer-aided knowledge sharing more effective (Berners-Lee & Hendler, 2001; Gerstein et al., 2007; Power et al., 1998; Uren et al., 2006). In fact, we already do this when we choose keywords from a categorized list provided by a journal publisher or conference website. When combined with lexicons and term dictionaries, such keyword selection can greatly increase search accuracy because the authors are best able to select which keywords most accurately reflect the content of the publication (Gil-Leiva & Alonso-Arroyo, 2007). For example, it is clear that controlled terminologies such as the PACS codes in physics, the JEL classification in economics, and the MeSH vocabulary in biomedical science, increase the accuracy of knowledge resource retrieval in those fields. Structuring a set of keywords in a controlled vocabulary can yield even better results (Kajikawa et al., 2006).

Several technologies being developed for the Semantic Web could help to realize an interactive knowledge sharing platform that would act as a forum for exchanging and integrating different forms of domain knowledge related to specific issues of sustainability science (Berners-Lee & Hendler, 2001; Wang et al., 2005; Hendler, 2005; McGuiness, 2004). In particular, controlled vocabularies enriched with explicit representation of semantic relationships between concepts in the vocabulary, known as “ontologies”, are expected to play an important role in realizing a more intelligent role of computers in handling knowledge sharing and discovery in the Web (Hess & Schliedera, 2006; Kraines et al., 2005). However, in order to make it possible for automated computer technologies to add value to scientific knowledge publications through knowledge mining and integration, we must address the problem of semantic interoperability (Heflin & Hendler, 2000).

The effectiveness of a computer to interoperate semantically with knowledge from different sources depends on what level of “understanding” the computer program can achieve of the knowledge (Wang et al., 2004; Halaschek-Wiener & Kolovski, 2008; Feigenbaum et al., 2007). This understanding must be conveyed through the semantics of the descriptive representations for each knowledge resource (Di Noia et al., 2007; Uschold et al., 2003). In particular, we know that computer reasoning algorithms are not good at “understanding” the semantics of statements written in natural language – natural language is simply too complex for present-day artificial intelligence (Natarajan et al., 2005; Hunter & Cohen, 2006; Basili et al., 2007). Consequently, if the computer-understandable descriptive representations of knowledge resources must be generated by natural language processing (NLP) techniques from natural language text, only the most basic semantics can be extracted accurately (Rzhetsky et al., 2008). However, if humans are empowered to create the computer-understandable descriptive representations of their knowledge resources, we gain access to the full capacity of humans to make sense of natural language (Power et al., 1998; Donaldson et al., 2003). Furthermore, if the human creating the computer-understandable descriptive representation is the same human that created the original knowledge resource, we can circumvent many of the remaining issues related to interpretation of natural language text such as ambiguity and information loss (Mons et al., 2008; Gao et al., 2006; Warner, 2007; Jensen et al., 2006; Hunter et al., 2008).

In summary, if computer-understandable semantic statements describing expert scientific knowledge can be provided that are highly accurate and semantically rich, we could leverage advanced inference and reasoning technologies from the fields of artificial intelligence and the Semantic Web to enable computers to provide a wide range of knowledge processing services. As one example, we suggest that it should be possible to translate semantic statements created in one natural language (such as Japanese) to another (such as English), or even from one domain of knowledge (such as chemical engineering) to another (such as macroeconomics), with essentially no loss of accuracy, simply by using a relatively straightforward mapping of content models that are expressed in terms of ontologies for different knowledge domains (Bateman, 1990; Bateman et al., 2005; Dymetman, 2002; Kruijff-Korbayova & Kruijff, 1999; Kraines & Guo, 2009).

3 METHODS

Our method for developing a knowledge-sharing platform for sustainability science is based on our previous work to develop tools and knowledge models for constructing web-based networks of expert knowledge (Kraines et al., 2005). Specifically, we have been developing four basic types of components: formalized ontologies grounded in description logics for describing concepts of a particular domain of knowledge, a repository of semantic statements created using the ontology to describe expert knowledge in that domain in a computer-understandable form, a semantic matching algorithm based on logic and rule-based inference provided by commercially available computer reasoning software, and graphing tools for visualizing and analyzing the resulting knowledge network.

These technologies are being integrated into a web-based platform called EKOSS, for “Expert Knowledge Ontology-based Semantic Search” (Kraines et al., 2006a). The EKOSS system is intended to provide a platform for effective computer-mediated sharing and integration of expert knowledge resources. To meet this goal, EKOSS provides two basic types of knowledge services. On the knowledge sharing side, EKOSS provides a set of web interfaces that allow knowledge creators, who do not have any special expertise in ICTs or knowledge representation, to create their own semantic statements describing the knowledge resources that they want to share in a form that can be interpreted semantically, and thus “understood”, by computers. These interfaces have been described in other papers (e.g. Kraines et al., 2006a) and can be accessed at http://www.ekoss.org (Kraines et al., 2011).

On the knowledge integration side, EKOSS is equipped with computer algorithms that use a reasoning engine, such as RacerPro (Racer Systems GmbH & Co. KG, 2011), to establish the degree of semantic proximity between semantic statements of different knowledge resources or requirements (e.g. queries expressing particular information needs) based on a number of criteria, including similarity of logical structure at a semantic level. From these calculated degrees of semantic proximity, a knowledge network can be created where the nodes are the semantic statements describing particular knowledge resources and the links are the semantic relationships established between those statements by the reasoning engine.

We have argued in the previous section that the semantic statements describing knowledge resources must be authored by the creators of the knowledge resources themselves. However, we face a typical “chicken and the egg” problem here. In order to convince a researcher to take the time to create semantic statements describing his or her knowledge resources, we need to demonstrate the benefits of doing so to the researcher. But in order to demonstrate the benefits of the knowledge sharing system, we need a sufficient number of semantic statements in the system. To “jump-start” this cycle, we have been using the EKOSS system to create semantic statements describing knowledge resources of researchers in three domains: human life sciences, engineering failures, and sustainability science. Our work to apply EKOSS to the domains of human life sciences and engineering failures is described elsewhere (Kraines & Guo 2010; Kraines et al., 2010). Towards the construction of a knowledge sharing platform for sustainability science, we have developed an extension of the ontology for the domain of engineering failures that encompasses concepts related to sustainability science, and we have created a set of semantic statements in EKOSS using that ontology to describe research related to sustainability science. Preliminary results of this work are described in the next section.

4 RESULTS

4.1 Ontology for concepts of sustainability science

At the core of the EKOSS-based knowledge sharing system for sustainability science is a formalized ontology based on a description logic that provides a knowledge model for describing concepts from the domain of sustainability science. This ontology is an extended version of the SCINTENG ontology, some applications of which have been presented in previous work (Kraines et al., 2005; Kraines et al., 2006b; Kraines et al., 2007). The SCINTENG ontology in turn is based on ISO 15926, which is “an integrated data model that covers the information requirements of the full process plant life cycle and enables integration of process plant information with other enterprise information” (European Process Industries STEP Technical Liaison Executive, 2009; Batres et al., 2007). Here we describe the underlying knowledge model that we developed for coordinating the construction of the ontology and how this knowledge model enables the creation of computer-understandable semantic statements describing case studies and scenarios in sustainability science.

Currently, there is no widely accepted ontology based on formal logic for expressing knowledge related to the domains of engineering failures and sustainability science. Our hypothesis is that if humans used a formal knowledge representation language to author semantic statements that describe their own knowledge resources, computers could perform more intelligent calculation of semantic similarity between pairs of resources. To test this hypothesis, we have developed an ontology, called SCINTENG, in previous work. The SCINTENG ontology draws on conceptualizations in thermodynamics (Atkins & de Paula, 2002), transport phenomena (Bird et al., 1960), the periodic table of elements, failure knowledge (Hatamura, 2005), and environmental science (Yasui, 2009). The schematic diagram of the basic SCINTENG knowledge model shown in Figure 1 illustrates the main classes in the ontology – "activity", "event", "physical object", and "substance or material" – and the main types of relationships that are supported between the different classes. Note that the class "physical object" includes states of physical objects as well as whole life individuals. In addition, several other classes that play special roles in the knowledge model are also shown, including "class of activity", "actor" and "location".

The knowledge model represents processes occurring over time by using the activity-event model of ISO 15926 (Batres et al., 2007), which is an upper-level ontology formalized in a description logic and based on a four-dimensional conceptualization of the world that is capable of describing complex processes and life cycle issues related to engineering artifacts. From the ISO 15926 documentation, an activity is defined as "a [possible individual] that brings about change by causing the [event] that marks the [beginning], or the [event] that marks the [ending] of a [possible individual]." An event is defined as "a [possible individual] with zero extent in time...{that is} the temporal boundary of one or more [possible individual]s..." (European Process Industries STEP Technical Liaison Executive, 2009). Consequently, the temporal model of ISO 15926 is based on special types of possible individuals called "activities" that are bounded in time by a start event and an end event. Activities can belong to specific categories described by "class of activity", have locations given by subclasses of physical objects called "spatial locations", and have partitive relationships with other activities through the property "has subactivity".

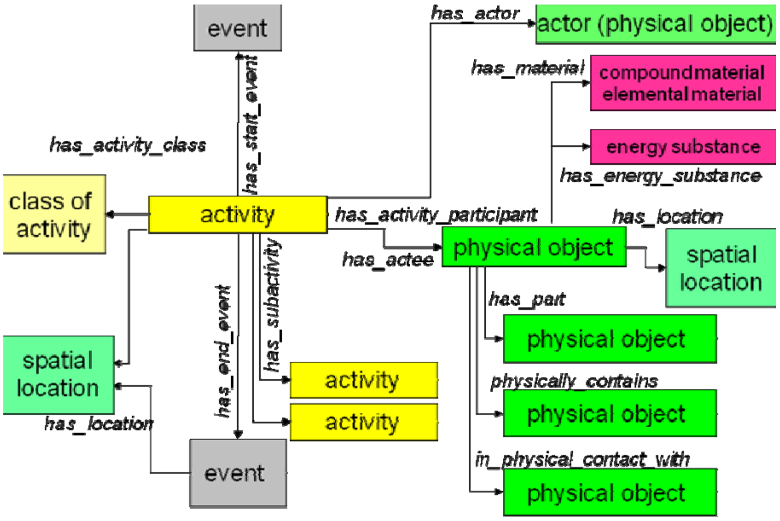

Figure 1. The knowledge model for the original version of SCINTENG. Major classes are shown in boxes, and major connecting properties are shown as directed arrows. The primary classes are activity (shown in yellow), physical object (shown in green), event (shown in gray), and substance (shown in pink). The labels of the arrows show the properties that can connect the origin class as a domain with the destination class as a range. Boxes with the same name are interchangeable, e.g. all activities can have a has_actee relationship with all physical objects.

In the SCINTENG ontology knowledge model, "physical objects" can participate in several kinds of active and passive ways with both activities and events. Compared to "activities", "physical objects" are considered to be invariant entities defined by the region of space that they occupy, although temporal bounds and changes of physical objects can also be described. In fact, in many cases a particular entity would be both a physical object and an activity, e.g. a storm or a forest fire. Finally, physical objects can have compositional and topological relationships with other physical objects, and they can have compositional relationships with materials or energy substances, which are considered to be sets of physical objects whose members share the property of being composed of the particular material. Physical objects are not required to have mass, so that a flame or an amount of electricity would be considered a physical object.

The SCINTENG ontology provides a framework for describing a wide range of phenomena and situational knowledge in engineering, material science, and sustainability science. We have also added a simple mechanism for providing a meta-layer to describe the analysis methods and tools used to study the situations and phenomena that are described. The classes for describing analysis methods and tools are defined as sub concepts of abstract object, which is disjoint from the concept tree containing activities and physical objects. This ensures that the analysis related classes are orthogonal to the main knowledge conceptualization. Several properties are provided to relate analysis methods and tools with activities and physical objects, including "has object of analysis” and “has concrete result of analysis”. An example is given in the next section.

As a part of a project for developing scenarios for Low-Carbon Societies, we began studying the applicability of the SCINTENG ontology to the domain of sustainability science. We realized that a third descriptive mechanism was needed to describe scenarios for achieving sustainability, such as the Japan Scenarios and Actions towards Low-Carbon Societies (hereafter the LCS scenarios) (2050 Japan Low-Carbon Society scenario team, 2008). Specifically, constructs were required to describe scenario-related concepts such as problem type, goal, alterative scenarios, and conditions for goal attainment.

Based on an analysis of the LCS scenarios, together with reference to other recent efforts to design knowledge conceptualizations for sustainability science (Kajikawa et al., 2008; Kumazawa et al., 2009; Hiramatsu et al., 2008), we developed a scenario description component for the SCINTENG ontology. The scenario version of the SCINTENG ontology introduces the following major classes: “scenario activity” as a subclass of “activity”, “scenario event” as a subclass of “event”, “goal type” and “problem type” as subclasses of “class of event”, “policy artifact” as a subclass of “artificial physical object”, and a variety of new event conditions for specifying requirements to achieve a certain target event. In addition, several new properties were introduced to relate both the new classes and classes from the original ontology in different ways that were considered important for describing scenarios. For example, the control mechanism classes are important subclasses of “class of activity” used to specify the economic, legal, technological and educational mechanisms that are used to control some human activity. New properties were introduced so that an activity or event can be described as creating a control mechanism, and a policy artifact can be described to embody a control mechanism. Control mechanisms can also be described as having implementing actors and targeted actors, similar to the ontology described by Kumazawa et al. (2009), and control mechanisms can be linked to the specific problem types or goal types that they are intended to address. In addition, a set of properties were included to make it possible to state that a scenario activity targets some situational activity, event, physical object, or even event condition. Finally, a set of properties were formulated to enable the description of convergent and divergent scenario paths by stating that an activity is either required for or sufficient for achieving a certain target event.

A view of how the three components of the revised SCINTENG ontology interrelate is shown in Figure 2, and details of the scenario conceptualization knowledge model are given in Figure 3.

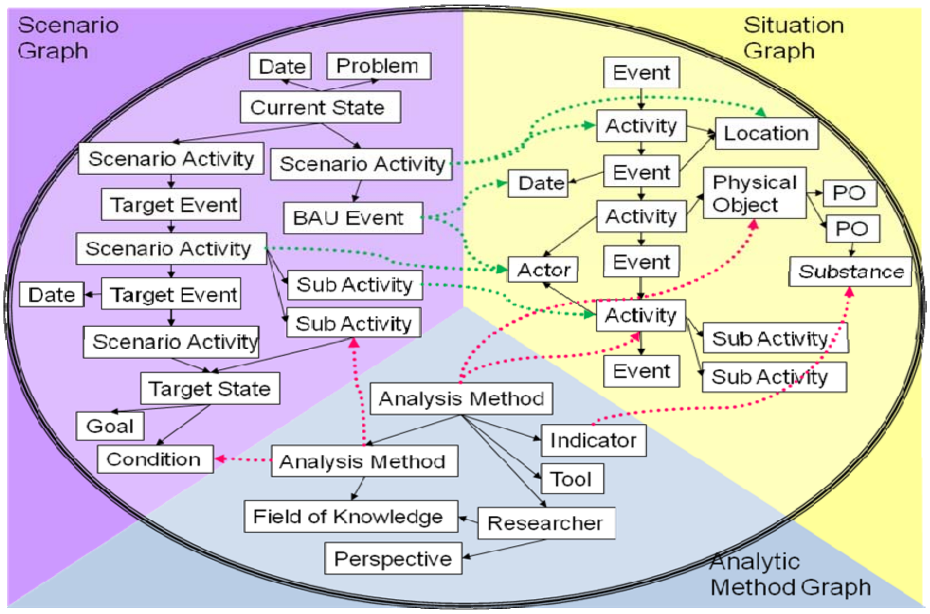

Figure 2. An illustration of how descriptions that are related to the three major conceptualizations of sustainability science – situations, scenarios, and analysis – are supported by the extended version of the SCINTENG ontology. Instances of classes from each of the conceptualization domains – scenario, situation, and analysis – are shown with boxes. Solid arrows show properties for relating concepts within each conceptualization, and colored dotted arrows show properties that relate concepts from different conceptualizations.

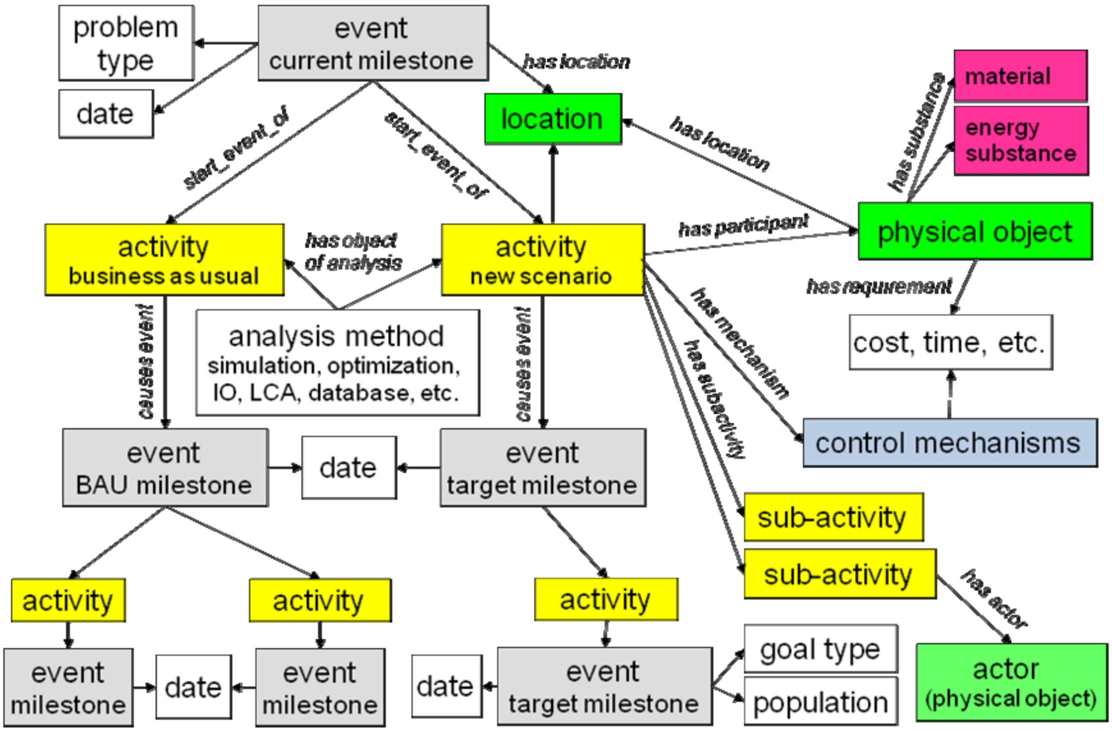

Figure 3. The knowledge model for the sustainability science version of SCINTENG. Symbol meanings are the same as in Figure 1. Major classes in the ontology are "activity" shown in yellow, "event" shown in gray, "object" shown in green, "substance" shown in pink and "mechanism" shown in light blue. In addition, there are several major abstract classes shown in white, such as date and analysis method. Activity is divided into human activities, natural activities, and scenario activities. Event is divided into normal events used to link human and natural activities, and scenario event used to link scenario activities. Objects include artificial and natural physical objects, energy objects, and non-substantial objects such as policy tools. Substances are divided into material substances that can be the substance of a natural or artificial physical object and energy substances that can be the substance of an energy object. Finally, mechanisms are a classification of activities that describe the manner in which the activity is controlled or carried out. The control mechanisms are classified as legal, economic, educational and technology, after Yasui (2009).

In the next section, we give several examples of how the extended SCINTENG ontology can be used to create computer-understandable semantic statements of expert knowledge in sustainability science.

4.2 Semantic statements for scenarios and technologies

The extended SCINTENG ontology provides a formal knowledge representation language for creating unambiguous, computer-understandable statements describing expert knowledge related to policy scenarios and individual technology options in the domain of sustainability science. We are working on creating a repository of semantic statements describing a wide range of expert knowledge in sustainability science. Towards that effort, we have created several sets of semantic statements using EKOSS. One set of 24 semantic statements has been created to describe proposals for research projects that have been funded by the AGS Promotion Office at the University of Tokyo between 2000 and 2005 (AGS Promotion Office at the University of Tokyo, 2009). Another set of over 100 semantic statements is based on papers published by individual researchers studying issues related to sustainability science and energy. The third set contains 7 semantic statements describing scenarios and technologies aimed at achieving a low-carbon society (2050 Japan Low-Carbon Society scenario team, 2008). Details for the semantic statements are summarized in Table 1. Note that the size of the triple store is quite large – even without including the results of inference along the class and property hierarchies in the ontology and the relationships that are implied by property symmetry and inverses, the triple store contains over 9000 triples to represent the class, label, and relationship information for the semantic statements.

Table 1. Statistics for the semantic statements from the domain of sustainability science

Type of resource	**# statements**	**avg # instances per statement**	**avg # properties per statement**
AGS proposals	24	23	28
Research papers	104	19	22
LCS scenarios	7	37	45
Total	135	21	24

Next we describe examples of each type of semantic statement and how they function to express the original knowledge content in a logical and unambiguous way that can be “understood” by computer algorithms for accomplishing “intelligent” tasks such as semantic matching and extraction of common semantic motifs from large sets of statements.

The first example is a semantic statement created to describe the proposed target of a research proposal that was accepted for funding by the AGS Promotion Office. The semantic statement is based on the following proposal abstract:

> *For realization of a sustainable global society, new technologies that are currently still unproved must penetrate successfully into the social infrastructure. Generally, when new technologies are introduced to a particular region, they complete and displace existing technologies, resulting in large-scale changes to production facilities, infrastructure, material transportation, societal standards, and other aspects of the region's industrial ecology. Rapid and "non-regret" introduction of new technologies requires consideration of the life cycle impacts of the system that is to use the new technology and the process for manufacturing the system from the research and development stage, in addition to consideration of the particular region characteristics. Furthermore, "trouble" or "failure" mechanisms that might occur at the manufacturing, operation, implementation, and dismantling stages must be considered. Development of a design support system for assessing design alternatives and failure prevention measures during the process of developing the new technology could help answer these issues. Using the budding energy technology "Solid Oxide Fuel Cell" as a concrete example of a new technology, we have studied the feasibility and potential effectiveness of a model-based technology design platform to support design of systems that could successfully penetrate into existing industrial ecologies.*

The graph view of the semantic statement created based on this abstract is shown in Figure 4.

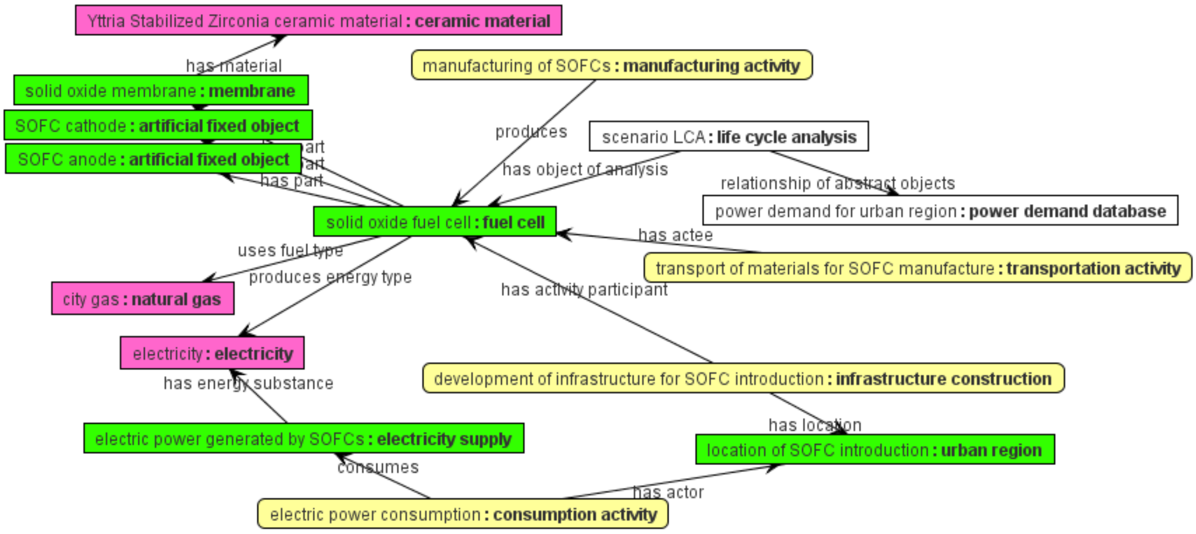

Figure 4. Graph view of semantic statement based on the abstract of the AGS-funded proposal entitled “Design Platform for the Penetration of Sustainable Technologies”. Instances of ontology classes describing entities from the text abstract are shown with boxes colored according to the schema in Figure 3 and containing the free text name of the instance followed by a colon and the name of the class. Properties specifying the relationship between the instances are shown with directed arrows labeled with the name of the property.

The semantic statement shown in Figure 4 can be rendered in English using a simple natural language generation algorithm that we have developed for the EKOSS system as shown in Figure 5 (Kraines & Guo, 2009). We have also developed natural language generation algorithms for Japanese and Chinese. Both the graph and the representation created by the natural language generator show clearly and unambiguously the specific manner in which the concepts that are addressed in the proposed research are related. For example, we know that the ceramic material “Yttria Stabilized Zirconia ceramic material” is the material of the membrane part of the fuel cell studied in the research. We can also see that the urban region is responsible for the consumption of the electricity supply produced by the fuel cell, which in turn consumes natural gas.

- **Solid oxide fuel cell** is a *fuel cell* that has part an *artificial fixed object* called **SOFC anode**, an *artificial fixed object* called **SOFC cathode**, and a *membrane* called **solid oxide membrane**, which is made of a *ceramic material* called **Yttria Stabilized Zirconia ceramic material**.
- The **solid oxide fuel cell** uses some *natural gas* called **city gas** and produces some *electricity*, which is the energy substance of an *electricity supply* called **electric power generated by SOFCs**.
- **Electric power consumption** is a *consumption activity* that consumes the **electric power generated by SOFCs** and that has actor an *urban region* called **location of SOFC introduction**.
- **Development of infrastructure for SOFC introduction** is an *infrastructure construction* that involves the **solid oxide fuel cell** and that is located in the **location of SOFC introduction**.
- **Scenario LCA** is a *life cycle analysis* that has object of analysis the **solid oxide fuel cell** and that is related to a *power demand database* called **power demand for urban region**.
- **Manufacturing of SOFCs** is a *manufacturing activity* that produces the **solid oxide fuel cell**.
- **Transport of materials for SOFC manufacture** is a *transportation activity* that involves the **solid oxide fuel cell**.

Figure 5. Natural language representation of the semantic statement shown in Figure 4. Instance names are shown in bold type, class names in italic type, and property names in green font. Domain instances, functioning as sentence subjects, are shown in blue, and range instances, functioning as sentence objects, are shown in red.

The second example, shown in Figure 6, is a semantic statement created to describe the research focus of a paper on a topic related to sustainability science.

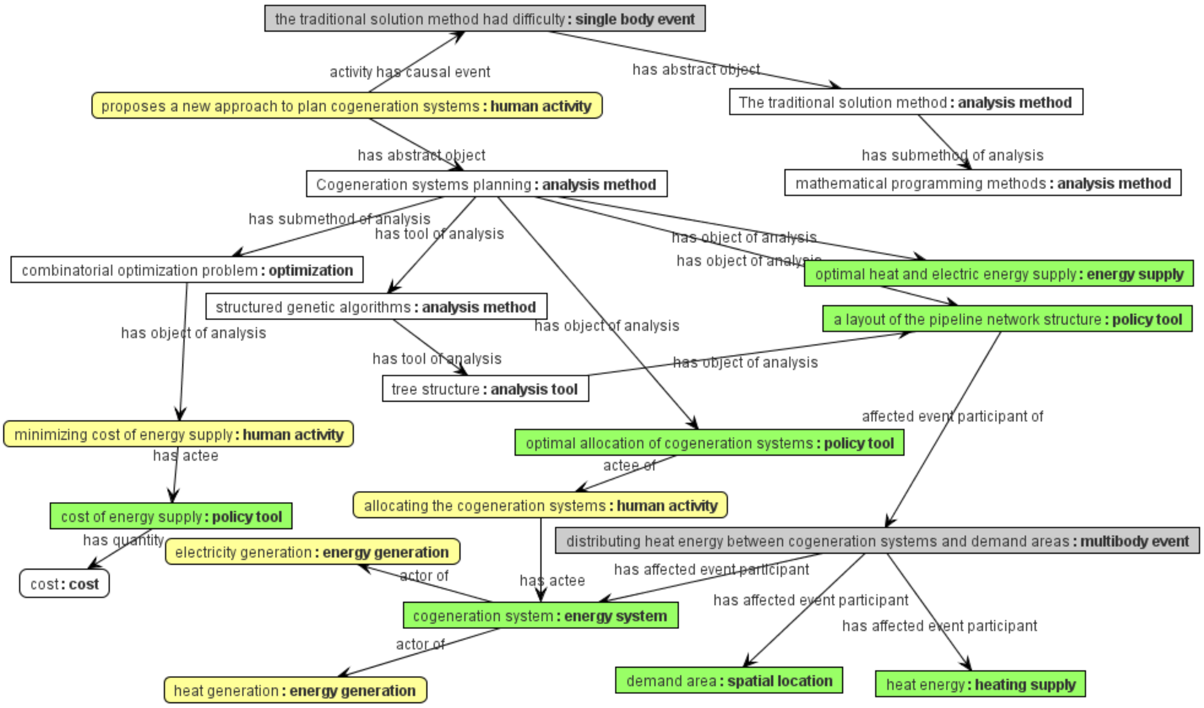

Figure 6. Graph view of semantic statement based on the abstract of the paper entitled “Cogeneration systems planning using structured genetic algorithms” by T. Harada and S. Mori (1997 Electrical Engineering in Japan Volume 119 Issue 2, Pages 26 – 35 (English translation of *Denki Gakkai Ronbunshi*)). Symbol meanings are the same as in Figure 4.

This example shows how instances of “analysis method” and “analysis tool” are used to specify what aspects of the situation described were studied by the researchers who wrote the paper. For example, we see that an analysis tool called “tree structure” was used to study a policy tool that was affected by an event involving a specified energy system and heating supply for a specified spatial location.

The third example, shown in Figure 7, is a semantic statement created to describe one of the LCS scenarios: “Next Generation Fuels: Biomass” (2050 Japan Low-Carbon Society scenario team, 2008).

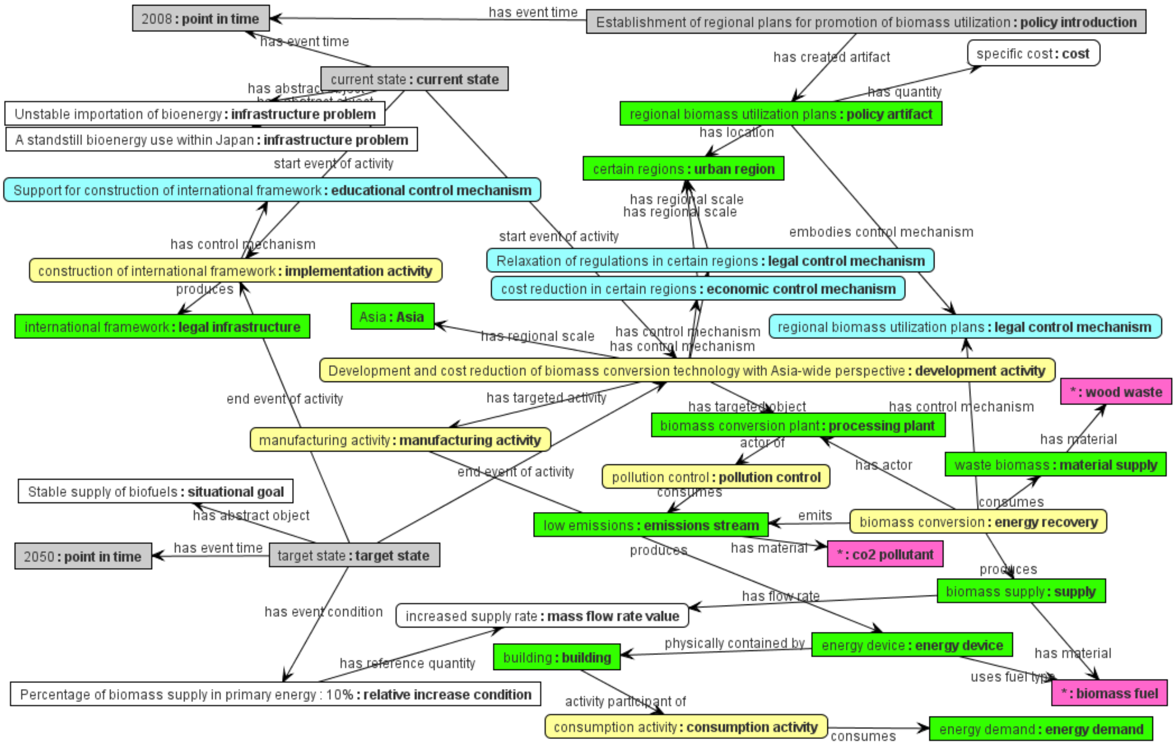

Figure 7. Graph view of semantic statement based on the abstract of the LCS scenario entitled “Next Generation Fuels: Biomass”. Symbol meanings are the same as in Figure 4.

This is a much more complex graph that involves both a scenario description and details of the phenomena and technologies that are targeted by the scenario. The scenario description begins with an instance of “current state” characterized by an event time and two problem types. The current state, which co-occurs with a policy introduction event, proceeds through two scenario activities, an implementation activity and a development activity, to a target state that is characterized by an event time, a goal type, and an event condition. The development activity involves several human activities, such as a manufacturing activity and an energy recovery activity, which in turn involve several physical objects, such as material supplies and processing plants. Although the graph shown in Figure 7 may be too confusing for human understanding, remember that the role of the semantic statement is not to be human understandable but to be computer-understandable. Together, these class instances and property relationships specify a complex scenario in a form that can be interpreted by a computer clearly and unambiguously, making it possible for a computer to precisely interpret and reason with the semantics of the description.

4.3 Examples of semantic matchmaking using EKOSS

Semantic statements can be used to calculate the similarity between the resources that they represent at the level of semantic relationships. As an example, we describe the results of matching the graphs in Figures 4 and 6 semantically with the graph in Figure 7. Details on the semantic graph matching techniques are given in (Guo & Kraines, 2008b).

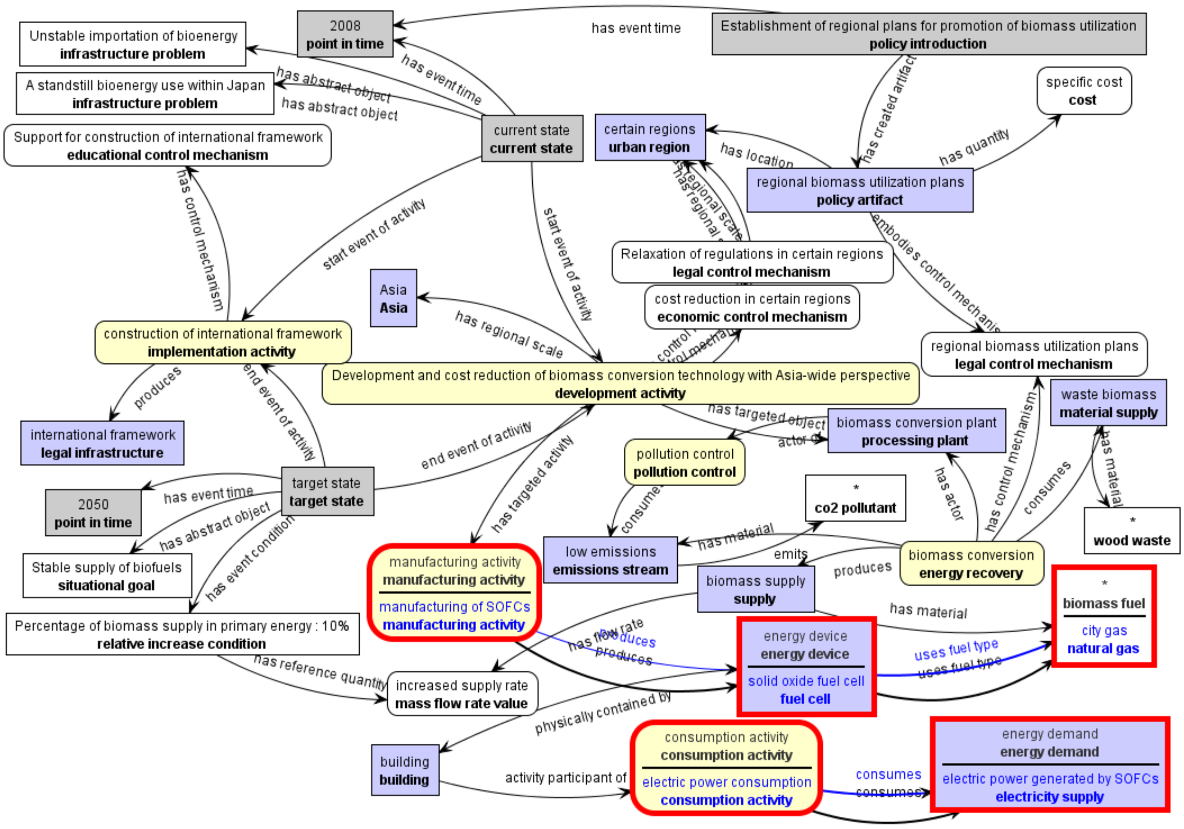

Figure 8. Graph view of the match of the search semantic statement describing the LCS scenario “Next Generation Fuels: Biomass” with the target semantic statement of the research proposal “Design Platform for the Penetration of Sustainable Technologies”. The full search statement is shown in black, and the matching instances and properties from the target statement are shown in blue. Boxes with thin black borders represent instances in the search statement, with the instance label on the top line and the class name on the bottom line. The boxes containing the matching instances from the search statement and the target statement are highlighted with thick red borders. Other symbol meanings are the same as in Figure 4.

Figures 8 and 9 show the results of using the semantic statement for the LCS scenario “Next Generation Fuels: Biomass” as the search statement to match with the semantic statements shown in Figures 4 and 6. From Figure 8, we see that the knowledge expressed in the research proposal could relate to the manufacturing of the energy device for meeting the energy demand of the consumption activities for the buildings in the target state. From Figure 9, we see that the knowledge contained in the research paper could be related to the development of the processing plant together with the related energy recovery and pollution control processes. In addition, there appears to be knowledge relating to the cost of the policy tools required for realization of the target state in the scenario described.

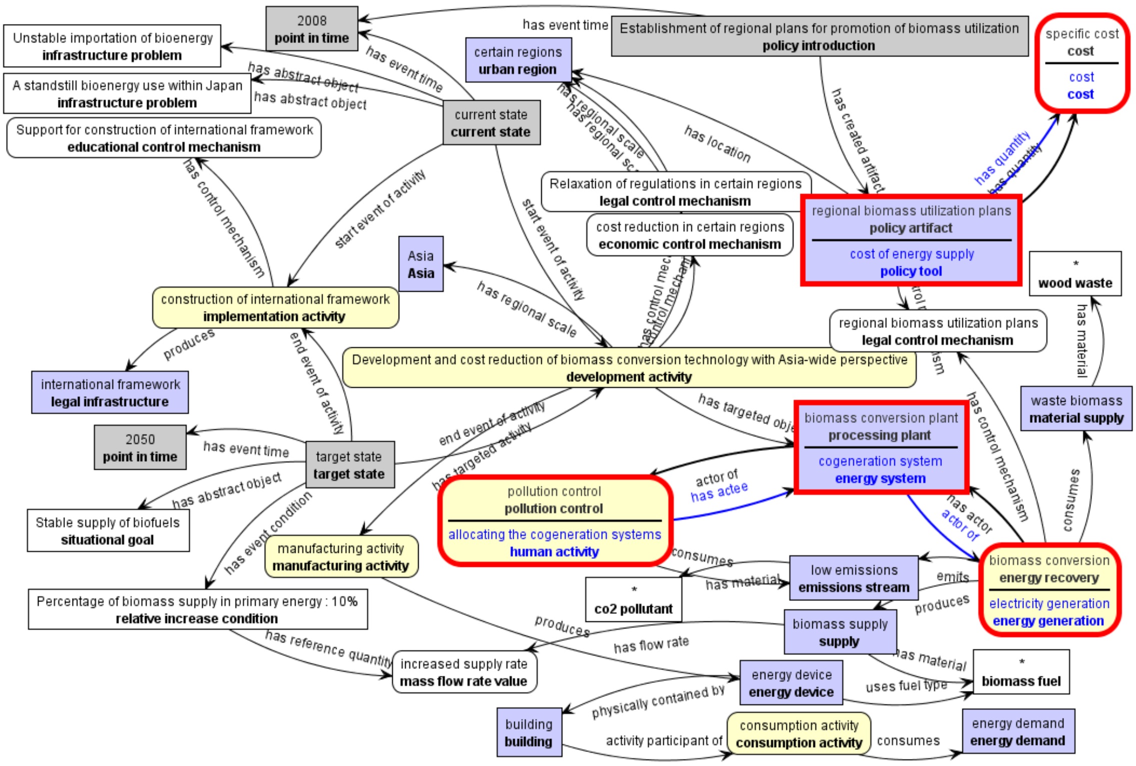

Figure 9. Graph view of the match of the search semantic statement describing the LCS scenario “Next Generation Fuels: Biomass” with the target semantic statement of the research paper “Cogeneration systems planning using structured genetic algorithms”. Symbol meanings are the same as in Figure 8.

5 CONCLUSIONS AND FUTURE DIRECTIONS OF RESEARCH

The EKOSS-based sustainability science knowledge sharing system is envisaged to provide the following main facilities.

1) Establishment of a semantic search engine that anyone can use to find researchers doing work that is related to a particular search condition expressing specific relationships between concepts of interest, not just a “bag of words” containing a list of disconnected concepts
2) Creation of a map of the domain of sustainability science, based on semantic similarity rather than citation or coauthor links, that could help researchers to visualize the state of the art in the domain as well as to determine how their research work is related semantically rather than bibliographically to the specific research studies of other experts in the domain
3) Extraction of serendipitous relationships between researchers based on evaluation of similarities and differences in semantic statements using logic and rule-based inference, and synthesis of new knowledge through logical induction, e.g., by mining common semantic structures from large repositories of semantic statements (Guo & Kraines, 2008a)

The EKOSS (Expert Knowledge Ontology-based Semantic Search) system provides the enabling technology for realizing this knowledge sharing system. Our work in developing the EKOSS system for assisting researchers in sharing, discovering, and integrating expert knowledge consists of two main parts: 1) implementation of user interface tools for assisting knowledge experts to easily and accurately create computer-interpretable semantic statements describing their knowledge resources, and 2) research activities aimed towards demonstrating the value that is added by the creation of these statements. In particular, a major component of the second research activity has been the development of matching algorithms, graphing tools and other software programs for visualizing and analyzing the semantic relationships between the semantic statements created through the EKOSS tools emerging from the first research activity.

The long term goal of our research is to create a comprehensive network of expert knowledge related to sustainability science that is easily accessible by users ranging from academic researchers to concerned individuals in society with no specific scientific training. In contrast to existing knowledge networking systems for academics such as VIVO, our approach uses formal knowledge representation languages to describe knowledge resources in a computer-understandable way, so that computers can infer semantic similarities between resources based on reasoning about predicate relationships. Specifically, this knowledge network would provide a platform for exploratory analysis of the interrelationships and logical conclusions that can be drawn from the semantic statements that have been created through the EKOSS system. This analysis is expected to help in the identification of critical connections between knowledge from different studies and possibly the establishment of new knowledge through integrative analysis techniques such as inductive logical reasoning based on graph mining (Guo & Kraines, 2008a). Through the use of the components that we are developing, we aim to create a web-based platform providing access to this network for a wide variety of applications. Furthermore, we hope that by interconnecting the resulting knowledge network for sustainability science with the other knowledge networks we are developing for engineering failure mechanisms and for life sciences, it will be possible to make useful scientific discoveries that integrate the knowledge from the three domains. In this way, our approach could function as an extension of literature-based scientific discovery (Swanson, 1990).

Finally, by linking the semantic search engine and knowledge network services we are developing with the existing infrastructure for dissemination of scientific knowledge, we hope to support the efforts to realize a more effective transfer of knowledge from academic fields to society and foster active relationships between researchers at the university and major actors in both the private and public sectors of society. The work presented here represents a foundational step towards realizing this goal.

In order to realize these long-term goals, we are pursuing research efforts and tool development in the following areas. First, we are continuing to refine the EKOSS tools for authoring semantic statements. In particular, we are working to integrate natural language processing (NLP) tools into the authoring process. NLP tools can be used to help authors identify the classes and properties in the ontology that best express their expert knowledge or information requirements. Also, we are examining the possibility of using a set of NLP tools in combination with machine learning techniques to provide semi-automatic authoring of semantic statements that would subsequently be checked and refined by the users.

In parallel to this work on developing the semantic statement authoring tools, we will continue to develop semantic matching algorithms and to use the semantic statements that we have created to date to test the ability of those algorithms to find interesting relationships that would not be apparent without the level of semantics available for computer processing in our system. For example, we are developing a method for comparing the precision and recall of pair-wise matching results within a set of semantic statements using only classes (discarding all property information), using classes and properties without any inference, and using classes and properties with logic and/or rule-based inference. The hypothesis is that matching only with class information will yield high recall but low precision (many false positives), matching using property information but no inference will yield high precision but low recall (many false negatives), and matching using property information together with inference will yield the best combination of precision and recall. Preliminary calculations indicate that this hypothesis is supported (Kraines & Guo, forthcoming).

Finally, we will continue to add to the repository of semantic statements describing knowledge resources related to sustainability with the goal of achieving a content of 1000 entries, which we believe is the “critical mass” that is needed to demonstrate the effectiveness of the ontology-based knowledge sharing approach we are developing and start up the virtuous cycle of community authoring of semantic statements (Mons et al., 2008).

6 ACKNOWLEDGEMENTS

We thank Yutaka Nakamura and Eri Aoki for their assistance in creating the semantic statements and Matthew West for helping us to develop the SCINTENG ontology. This work was funded by the Global Environment Research Fund of the Japanese Ministry of the Environment, S-3-3: “Japan Low Carbon Society Scenarios toward 2050: Scenario development and its implication for policy measures, effects of introducing countermeasures for carbon dioxide emission reduction in urban areas” and by the Alliance for Global Sustainability Promotion Office at University of Tokyo.

7 REFERENCES

2050 Japan Low-Carbon Society scenario team (2008) Japan Scenarios and Actions towards Low-Carbon Societies (LCSs). Retrieved April 27, 2009 from the World Wide Web: http://2050.nies.go.jp/material/2050_LCS_Scenarios_Actions_English_080715.pdf

Abbott, R. (2004) Subjectivity as a concern for information science: a Popperian perspective. *Journal of Information Science* 30 (2), 95-l06.

AGS Promotion Office at the University of Tokyo. Research Projects. Retrieved April 27, 2009 from the World Wide Web: http://en.ags.dir.u-tokyo.ac.jp/research

Atkins P. W. & de Paula J. (2002) Atkins' Physical Chemistry 7th edition. Oxford University Press, Oxford.

Basili, R., de Cao, D., & Giannone, C. (2007) Ontological modeling for Interactive Question Answering. *Proc. OnToContent 2007* (pp. 544-553). Algarve, Portugal.

Bateman J.A., Kruijff-Korbayova I., & Kruijff G.-J. (2005) Multilingual resource sharing across both related an unrelated languages. *J. Research on Language and Computation* 3, 191-219.

Bateman J.A. (1990) Upper modelling: a general organization of knowledge for natural language processing. *USC/Information Sciences Institute.*

Batres, R., West, M., Leal, D., Price, D., Masaki, K., Shimada, Y., Fuchino, T., & Naka, Y. (2007) An upper ontology based on ISO 15926. *Computers & Chemical Engineering* 31, 519-534.

Berners-Lee, T., & Hendler, J. (2001) Publishing on the Semantic Web. *Nature* 410, 1023-1024.

Bird RB, Stewart WE, Lightfoot EN (1960) Transport Phenomena. Wiley, New York.

Borner, K., Chen, C., & Boyack, K. W. (2003) Visualizing Knowledge Domains. *Annual Review of Information Science and Technology* 37, 179-255.

Cahlik, T. (2000) Comparison of the maps of science. *Scientometrics* 49, 73-87.

Clark, W.C. (2007) Sustainability Science: A room of its own. *PNAS* 104(6), 737-1738.

Dymetman M. (2002) Text authoring, knowledge acquisition and description logics. *Proc 19th Intl Conf on Computational Linguistics,* Vol. 1.

Di Noia, T., Di Sciascio, E., & Donini, F.M. (2007) Semantic Matchmaking as Non-Monotonic Reasoning: A Description Logic Approach. *J. Artificial Intelligence Research* 29, 307.

Donaldson, I., Martin, J., de Bruijn, B., Wolting, C., Lay, V., Tuekam, B., Zhang, S., Baskin, B., Bader, G.D., Michalickova, K., Pawson, T., & Hogue, C.W.V. (2003) PreBIND and Textomy: mining the biomedical literature for protein-protein interactions using a support vector machine. *BMC Bioinformatics* 4, 11.

European Process Industries STEP Technical Liaison Executive: EPISTLE Core Model. Retrieved April 27, 2009 from the World Wide Web: http://www.btinternet.com/~Chris.Angus/epistle/specifications/ecm.html

Feigenbaum, L., Herman, I., Hongsermeier, T., Neumann, E., & Stephen, S. (2007) The Semantic Web in Action, *Scientific American* December 2007.

Fukushima, Y., Shimada, M., Kraines, S.B., Hirao, M., & Koyama, M. (2004) Scenarios of solid oxide fuel cell introduction into Japanese society. *Journal of Power Sources* 130(1-2), 327–339.

Gao, Y., Kinoshita, J., Wu, E., Miller, E., Lee, R., Seaborne, A., Cayzer, S., & Clark, T. (2006) SWAN: A distributed knowledge infrastructure for Alzheimer disease research. *Web Semantics: Science, Services and*

Agents on the World Wide Web 4(3), 222-228.

Garfield, E. (2001) A retrospective and prospective view of information retrieval and artificial intelligence in the 21st century. *JASIST* 52 (1), 18-21.

Gerstein, M., Seringhaus, M. & Fields, S. (2007) Structured digital abstract makes text mining easy. *Nature* 447, 142.

Gewin V. (2009) Networking in VIVO. *Nature* 462, 123 | 10.1038/nj7269-123a.

Gil-Leiva, I., & Alonso-Arroyo, A. (2007) Keywords given by authors of scientific articles in database descriptors. *JASIST* 58(8), 1175-1187.

Gunter, D. W. (2009) Semantic Search. *American Society for Information Science and Technology / ASIS&T Bulletin*, October/November 2009.

Guo, W., & Kraines, S. B. (2008a) Mining Common Semantic Patterns from Descriptions of Failure Knowledge. *Proceedings of MLG2008, July 4-5, 2008,* Helsinki, Finland.

Guo, W., & Kraines, S. B. (2008b) Explicit Scientific Knowledge Comparison Based on Semantic Description Matching. *Annual meeting of the ASIST 2008,* Columbus, Ohio.

Halaschek-Wiener, C., & Kolovski, V. (2008) Syndication on the Web using a description logic approach. *Web Semantics: Science, Services and Agents on the World Wide Web* 6(3), 171-190.

Hatamura, Y. (2005) Structure and expression of failure knowledge database. Retrieved April 27, 2009 from the World Wide Web: http://shippai.jst.go.jp/en/Contents?fn=1&id=GE0704.

Heflin, J. & Hendler, J. (2000) Semantic Interoperability on the Web. *Proceedings of Extreme Markup Languages 2000* (pp. 111-120) Graphic Communications Association, Alexandria, VA.

Hendler, J.A. (2005) Knowledge Is Power: A View from the Semantic Web. *AI Magazine* 26(4), 76-84.

Hess, C., & Schliedera, C. (2006) Ontology-based verification of core model conformity in conceptual modeling. *Computers, Environment and Urban Systems* 30(5), 543-561.

Hiramatsu, A., Mimura, N., & Sumi, A. (2008) A mapping of global warming research based on IPCC AR4. *Sustainability Science* 3, 201–213.

Hunter, L., & Cohen, K.B. (2006) Biomedical language processing: what's beyond PubMed? *Mol Cell.* 21(5), 589-94.

Hunter, L., Lu, Z., Firby, J., Baumgartner Jr., W.A., Johnson, H.L., Ogren, P.V., & Cohen, K.B. (2008) OpenDMAP: An open source, ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interactions and cell-type-specific gene expression. *BMC Bioinformatics* 9, 78.

Jensen, L.J., Saric, J., & Bork, P. (2006) Literature mining for the biologist: from information retrieval to biological discovery. *Nature Reviews Genetics* 7, 119-129.

Kajikawa, Y. (2008) Research core and framework of sustainability science. *Sustainability Science* 3, 215–239.

Kajikawa, Y., Abe, K., & Noda, S. (2006) Filling the gap between researchers studying different materials and different methods: a proposal for structured keywords. *Journal of Information Science* 32, 511-524.

Kajikawa, Y., Ohno, J., Takeda, Y., Matsushima, K., & Komiyama, H. (2007) Creating an academic landscape of sustainability science: an analysis of the citation network. *Sustainability Science* 2(2), 221-231.

Kostoff, R.N. (2005) Encouraging discovery and innovation. *Science*, 309(5732):245-6.

Kostoff, R.N. (2002) Overcoming Specialization. *BioScience,* 52(10): 937-941(5).

Kraines, S. B., Guo, W., Kemper, B., & Nakamura, Y. EKOSS: The Expert Knowledge Ontology-based Semantic Search System. Retrieved Jan 18, 2011 from http://www.ekoss.org

Kraines, S.B., Batres, R., Koyama, M., Wallace, D.R. & Komiyama, H. (2005) Internet-based integrated environmental assessment: using ontologies to share computational models. *J. Industrial Ecology* 9, 31-50.

Kraines, S. B., Guo, W., Kemper, B., & Nakamura, Y. (2006a) EKOSS: A knowledge-user centered approach to knowledge sharing, discovery, and integration on the Semantic Web. *Proc of the 2006 Intl Semantic Web Conf LNCS* 4273, 833-846.

Kraines, S. B. & Guo, W. (in preparation) Matching computer understandable knowledge descriptions of sustainability science research.

Kraines, S.B, Koyama, M., & Batres, R. (2006b) An emergent simulation modeling approach for discovery of knowledge on phenomena in chemical systems. *Journal of Chemical Engineering of Japan.: Society of Chemical Engineers* 39, 1010-1027.

Kraines, S.B., Batres, R., Kemper, B., Koyama, M., & Wolowski, J. (2007) Internet-based integrated environmental assessment: semantic searching based on ontologies and agent systems for knowledge. *Journal of Industrial Ecology* 10, 1-24.

Kraines, S. B. & Guo, W. (2009) Using human authored Description Logics ABoxes as concept models for Natural Language Generation. *Annual meeting of the ASIST 2009,*Vancouver, Canada.

Kraines, S. B. & Guo, W. (2010) Supporting reuse of knowledge of failures through ontology-based semantic search. *Proceedings of 2nd International Conference on Knowledge Management and Information Science 2010, KMIS 2010,* 26 October 2010, Valencia, Spain.

Kraines, S. B., Guo, W., Hoshiyama, D., Mizutani, H., Takagi, T. (2010) Generating literature-based knowledge discoveries in life sciences using relationship associations. *Proceedings of 2nd International Conference on Knowledge Discovery and Information Retrieval 2010, KDIR 2010,* 26 October 2010, Valencia, Spain.

Kruijff-Korbayová I. & Kruijff G.-J. M. (1999) Text Structuring in a Multilingual System for Generation of Instructions. In: V. Matousek et al. (Eds.): TSD'99, LNAI 1692 , 89-94. Springer-Verlag.

Kumazawa, T., Saito, O., Kozaki, K., Matsui, T., & Mizoguchi, R. (2009) Toward knowledge structuring of sustainability science based on ontology engineering. *Sustainability Science* 4(1), 99-116.

McGuinness, D.L. (2004) Question Answering on the Semantic Web. *IEEE Intelligent Systems* 19(1), 82-85.

Mons, B., Ashburner, M., Chichester, C., van Mulligen, E., Weeber, M., den Dunnen, J., van Ommen, G.J., Musen, M., Cockerill, M., Hermjakob, H., Mons, A., Packer, A., Pacheco, R., Lewis, S., Berkeley, A., Melton, W., Barris, N., Wales, J., Meijssen, G., Moeller, E., Roes, P.J., Borner, K., & Bairoch, A. (2008): Calling on a million minds for community annotation in WikiProteins. *Genome Biol.* 9(5), R89.

Natarajan, J., Berrar, D., Hack, C.J., & Dublitzky, W. (2005) Knowledge discovery in biology and biotechnology texts: A review of techniques, evaluation strategies, and applications. *Critical Reviews in Biotechnology* 25, 31-52.

Neumann, E., & Prusak, L. (2007) Knowledge networks in the age of the Semantic Web. *Briefings in Bioinfomatics* 8 (3), 141-149.

Power, R., Scott, S., & Evans, R. (1998) What You See Is What You Meant: direct knowledge editing with natural language feedback. ECAI 98. 13th European Conference on Artificial Intelligence. Edited by Henri

Prade. Published in 1998 by John Wiley & Sons, Ltd. Retrieved April 27, 2009 from the World Wide Web: http://mcs.open.ac.uk/rp3242/papers/ecai98.pdf.

Racer Systems GmbH & Co. KG.: World Wide Web site. Retrieved Jan 18, 2011 from the World Wide Web: http://www.racersystems.com

Rzhetsky, A., Seringhaus, M., & Gerstein, M. (2008) Seeking a new biology through text mining. *Cell* 134, 9-13.

Seringhaus, M. & Gerstein, M. (2008) Manually structured digital abstracts: a scaffold for automatic text mining. *FEBS Lett* 582, 1170.

Swanson, D.R. (1990) Medical literature as a potential source of new knowledge. *Bull. Med. Libr. Assoc.* 78, 29-37.

Takeuchi, K., & Komiyama, H. (2006) Sustainability science: building a new discipline. *Sustainability Science* 1(1), 1-6.

Tang, J., Zhang, J., Zhang, D., Yao, L., Zhu, C., & Li, J.-Z. (2007) ArnetMiner: An Expertise Oriented Search System for Web Community. *Semantic Web Challenge'2007*.

Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., & Ciravegna, F. (2006) Semantic annotation for knowledge management: requirements and a survey of the state of the art. *Web Semantics: Science, Services and Agents on the World Wide Web* 4 (1), 14-28.

Uschold, M., Clark, P., Dickey, F., Fung, C., Smith, S., Uczekaj, S., Wilke, M., Bechhofer, S., & Horrocks I. (2003) A semantic infosphere. *Proc. Intl. Semantic Web Conf,* Springer, pp. 882-896.

Wang, X., Gorlitsky, R., & Almeida, J. S. (2005) From XML to RDF: how semantic web technologies will change the design of ‘omic’ standards. *Nature Biotechnology* 23 (9), 1099-1103.

Wang, J., Jin, B., & Li, J. (2004) An Ontology-Based Publish/Subscribe System. *Proc. 5th ACM/IFIP/USENIX Intl Conf on Middleware* (pp. 232-253) Toronto, Canada.

Warner, J. (2007) Description and search labor for information retrieval. *JASIST* 58(12), 1783-1790.

Weikum, G., Kasneci, G., Ramanath, M., Suchanek, F. (2009) Database and Information-retrieval Methods for Knowledge Discovery. Communications of the ACM, 4, 56-64.

Yasui, I. (2009) Personal communication

(Article history: Received 25 January 2010, Accepted 14 December 2010, Available online 13 January 2011)