SUPPORTING REUSE OF KNOWLEDGE OF FAILURES THROUGH ONTOLOGY-BASED SEMANTIC SEARCH

Steven B. Kraines and Weisen Guo
Science Integration Program (Human), Department of Frontier Sciences and Science Integration Division of Project Coordination, The University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa 277-8568 Chiba, Japan

Keywords: Knowledge representation, Ontology, Semantic Search, Domain knowledge, Knowledge reuse, Failure knowledge.

Abstract: In order to increase the effectiveness of sharing and reusing knowledge about failures, we have applied the expert knowledge ontology-based semantic search system, EKOSS, to the "failure knowledge database". An ontology based on description logics is used as a formalized knowledge representation language for creating semantic statements describing 212 of the JST failure cases. Using the EKOSS reasoner, the similarity between a search statement giving the conditions of a new project or project design and the semantic statements describing failure cases can be quantified by inferring specific semantic relationships between entities involved in the case. The corpus of semantic statements is described, and results of applying the EKOSS semantic search to the semantic statements are analyzed. Finally, the effectiveness of the SCINTENG ontology for expressing the underlying failure mechanisms of the cases is discussed.

1 INTRODUCTION

Civil engineering projects and other applications of technology have become increasingly knowledge intensive. Institutions involved in construction, operation, and/or disposal of engineering artifacts such as buildings, bridges and airplanes, are held responsible for wide-ranging safety and environmental issues. Failures to meet promised standards can seriously impact public perception of the institution.

Many failures that plague institutions today could have been avoided if the knowledge of similar previous failures had been available and the relationship to the project concerned had been recognized (Tamura, 2003). The importance of "reusing" knowledge obtained from the analysis of actual failures and the elucidation of the mechanisms leading to the failures is widely recognized (Brown, 2007; Darlington and Booker, 2006; Wintle and Pargeter, 2005). One attempt to structure the knowledge from past failures so as to facilitate their reuse in the assessment of safety and potential failure mechanisms of new technology applications is the "failure knowledge database" (hereafter failure DB) (Hatamura et al., 2003a; Hatamura et al., 2003b; JST, 2010).

EKOSS (expert knowledge ontology-based semantic search) is a Web-based system for supporting the computer-aided sharing, discovery and integration of knowledge resources (Kraines et al., 2006a). Building on the premise that knowledge sharing and discovery could benefit greatly if human creators of knowledge resources created computer-interpretable descriptors of those resources (Gerstein et al., 2007; Marcondes, 2004), EKOSS provides knowledge creators with tools enabling them to easily construct computer-understandable descriptors, which we call "semantic statements", that describe their resources using ontologies based on description logics (DL) as knowledge representation languages. The semantic statement authoring tools incorporate several automatic support functions such as NLP of the source text in order to reduce the cognitive overhead for the human authors. EKOSS then uses the semantic statements to provide knowledge-intensive services such as semantic search and knowledge mining.

Semantic statements that are authored directly by the creators of the knowledge resources are expected to be more accurate and semantically rich than descriptors that are generated automatically using NLP (Berners-Lee and Hendler, 2001; Blake and Rendall, 2006; Buckingham Shum et al., 2007). In particular, a computer-mediated searching and matching system having access to EKOSS semantic statements could use logical and rule-based inference to provide

B. Kraines S. and Guo W..
SUPPORTING REUSE OF KNOWLEDGE OF FAILURES THROUGH ONTOLOGY-BASED SEMANTIC SEARCH.
DOI: 10.5220/0003068001640169
In *Proceedings of the International Conference on Knowledge Management and Information Sharing* (KMIS-2010), pages 164-169
ISBN: 978-989-8425-30-0

knowledge services that would not be possible with descriptors generated by NLP or written by humans as statements that are not formalized in a computable logic. Systems have been reported in the literature that attempt to test this hypothesis (Di Noia et al., 2007; Li and Horrocks, 2003; Uschold et al., 2003; Wang et al., 2004; Halaschek-Wiener and Kolovski, 2008). However, they do not support reasoning against DL expressions that include assertions of relationships between individuals described as instances of ontology classes, so it is not possible to create complex statements that involve, for example, relationships between different instances of the same class. Furthermore, most of these systems do not have user interfaces that are designed to enable users without information science backgrounds to create logically based computer-understandable semantic statements.

The EKOSS system has been designed to be integrated into the existing scientific knowledge publication process, so authoring the semantic statements would become just one step in the process of creating a scientific publication. The idea of incorporating creation of computer-interpretable descriptors in the process of publishing research articles has been raised before (Marcondes, 2004; Berners-Lee and Hendler, 2001; Rzhetsky et al., 2008; Ceol et al., 2008). However, previous attempts to design such a system have had limited success. The EKOSS system aims to overcome the limitations of these previous attempts by providing a solid foundation in computer-interpretable semantics together with the intuitive authoring tools (Kraines et al., 2006a).

Here, we describe the application of EKOSS to the failure DB. We first give the motivation for constructing the failure DB and for applying EKOSS to it. Next, we present a corpus of semantic statements that were created for 212 of the cases using EKOSS, and we show how the semantic richness of the statements can improve the accuracy of searching the failure DB for knowledge that could be reused in new engineering applications. Finally, we discuss the effectiveness of the ontology as a conceptualization for expressing the mechanisms behind technology failures and suggest some areas for improvement.

2 FAILURE KNOWLEDGE

In 2002, the Japan Science and Technology Agency (JST) started a project to create a database, the failure DB, containing case studies of major failures in science and technology (JST, 2010). One aim was to make the knowledge from the case studies reusable in new engineering applications to avoid repeat of past failures (Tamura, 2003; Hatamura et al., 2003a). Over five years, descriptions for over 1000 cases of major failures from around the world, including famous disasters such as the sinking of the Titanic, were prepared under the supervision of five experts on the analysis of failure mechanisms in chemical engineering, mechanical engineering, material science, and civil engineering. The descriptions were made available in both English and Japanese on the JST “shippai chishiki” website (JST, 2010).

The failure DB is structured via three “failure” Mandala, which classify types of causes, actions and results of failure processes (Hatamura et al., 2003b). For each failure case, key terms from each Mandala were chosen and arranged in a scenario that depicts the unfolding of the events surrounding the particular knowledge failure.

3 SEMANTIC STATEMENTS

Semantic statements were created for 212 of the failure cases, including the 100 “hyaku-sen” cases forming the core of the failure DB. The semantic statements were authored using a DL ontology, called SCINTENG, elements and applications of which have been presented elsewhere (Kraines et al., 2006b; Kraines et al., 2006c). By using a DL reasoner, knowledge sharing services can be provided based on inference at the level of semantic relationships, e.g. to identify failure cases that have similar causal mechanisms to new engineering failures.

Terms in the Mandala for “cause of failure” and for “action” were added to the SCINTENG ontology as subclasses of “class of human activity.” However, most terms in the Mandala for “result” are actually events or properties of physical objects. Therefore, we have included them in the SCINTENG ontology under the relevant parent classes of “disaster event” and “abnormality property.”

Using the EKOSS authoring tools, each semantic statement took one to four hours to create. The semantic statements were created using information given in the failure DB as follows:

1. Establish an event-activity chain that describes the course of the incident based mainly on the “sequence” and “cause” case descriptions.
2. Add physical objects, materials, locations, times, etc. by referring mainly to information in the “sequence” and “cause” case descriptions.
3. Refer to the “scenario” to add failure types to the activities.

The semantic statements contain an average of 35 instances and 45 properties each, for a total of 7455 instances and 9603 properties. Of the 1000 classes in the SCINTENG ontology, 507 were used at least once, and 135 were used 10 times or more. 38.6% of the instances are physical objects, 17.1% are activities, 17.2% are events, 8.8% are classes of activities, 6.9% are substances, and 11.4% are other classes such as quantities, units, and properties. Of the 195 properties in the ontology, 112 were used at least once, and 63 were used 10 or more times. Most of the properties are inverses or subproperties of “has activity participant” (15.6%), “has event participant” (14.7%), “has end event” (10.2%), “has start event” (8.8%), “has location” (8.1%), “has activity class” (6.9%), “has part” (6.0%), “has substance” (5.0%), or “physically contains” (5.2%).

4 QUERY MATCHING ANALYSIS

Direct measurement of the increase in precision and recall of knowledge searches that can be obtained by the EKOSS system requires enough semantic statements describing search targets such as case reports to make comparisons with conventional search engines. Because we only have a limited number of semantic statements, we have developed a different approach for assessing the ability of EKOSS to increase search precision and recall.

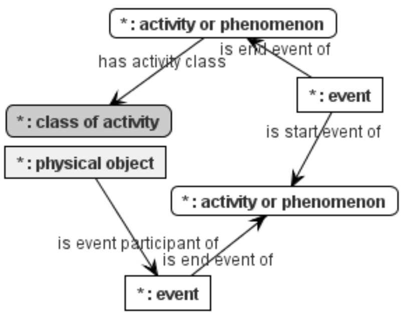

Figure 1: The general query for a search for knowledge describing some activity or phenomenon with a specified class of activity and an end event that is the start event of a second activity, which ends in a second event that has a participating physical object. Boxes show query variables and are labeled “class name”, colon, “instance label”. Physical objects and events are shown as grey and white square boxes. Activities and classes of activities are shown as white and grey rounded boxes. Arrows show properties asserted between pairs of query variables.

First, we created a general search query with six instance variables of top level classes and five properties from the SCINTENG ontology (Figure 1). This query matches with 88 of the 212 semantic statements, so we can consider it to be a commonly occurring semantic pattern. Next, we chose three semantic statements from the 88 matching the general query, and we used the mid-level classes of the ontology that occurred in at least 50 statements, shown in Table 1, to manually construct three search queries based on the general query that resulted in one-to-one matches (100% precision) with each of the selected semantic statements. The resulting “base queries” are shown in Tables 2 to 4. Instance variables are labeled by the name of the class followed by a capital letter to differentiate variables having the same class.

Table 1: Commonly occurring mid-level classes, the total count of instances of the class and all of its subclasses, and the major superclass from the knowledge model.

Class	Count	Top Class
artificial physical object	1108	physical object
single body event	475	event
urban region	413	spatial location
person	354	actor
disaster event	313	event
artifact actor	288	actor
machine artifact	288	physical object
termination event	267	event
multibody event	251	event
fluid object	248	physical object
operation	232	activity
reaction phenomenon	225	activity
industrial plant	135	physical object
chemical reaction	121	activity
mass transport	116	activity

Next, we removed the properties from the base queries to create queries containing just lists of mid-level classes, and we matched those queries against the 212 semantic statements. The number of additional matches that occur can be considered “false positives”, which indicate the increase in precision that was achieved by using the properties as a filter. Finally, we changed selected mid-level classes in the

Table 2: The base query for the case “Collapse of the Seongsu Bridge in Seoul, Korea,” hereafter “Seongsu Bridge.” Domain is the domain instance variable, range is the range instance range, and relation is the property connecting the two variables.

Domain	Relation	Range Instance
operationA	has class	human failureA
physical objectA	participant of	termination eventA
termination eventA	end of	mass transportA
termination eventB	start of	mass transportA
termination eventB	end of	operationA

base queries back to the corresponding top-level classes, keeping all of the properties this time. The new matches resulting from these operations are potential "true positives" that were missed by the base queries, which indicate the increase in recall achieved by using inference based on class subsumption. These operations to "relax" the query restrictions are shown in Table 5 together with the number of statements matching the query for each case.

Table 3: The base query for the case "Fire disaster in the Tokyo-Nagoya Nihon-zaka tunnel," hereafter "Nihon-zaka tunnel." See Table 2 for explanation of table columns.

Domain	Relation	Range Instance
artificial physical objectA	participant of	destruction eventA
destruction eventA	end of	chemical reactionA
eventA	start of	chemical reactionA
eventA	end of	operationA
operationA	has actor	artifact actorA
operationA	has class	human failureA

Table 4: The base query for the case "Leakage of primary coolant at Mihama Unit 2," hereafter "Leakage at Mihama." See Table 2 for explanation of table columns.

Domain	Relation	Range Instance
artificial physical objectA	has substance	elemental materialA
artificial physical objectA	participant of	reactionA
eventA	end of	human activityA
eventA	start of	reactionA
fluid objectA	participant of	single body eventA
human activityA	has class	human failureA
machine artifactA	has part	artificial physical objectA

5 DISCUSSION

Eliminating properties from the base queries added 6 to 13 matches (Table 5). This confirms that the one-to-one matches with the selected semantic statements did not result just from choosing rare combinations of classes. It also shows that the high precision of the base queries results at least in part from the assertion of properties between the instances.

Table 5: Results of the query relaxation analysis. "Relaxation Change" is the specific change made to the base query: ">>" indicates a class change, "HA" is "human activity" and "PO" is physical object. Labels "Rx" indicate that the change for the entry having the same label in the "Label" column was made, e.g. "R2+R3" means that "operation" was changed to "human activity" and "mass transport" was changed to "activity". "Num" is the number of matching cases, which includes the selected semantic statement, so there is always at least one match.

Case	Relaxation Change	Num	Label
Seongsu Bridge	remove all properties	13	-
	operation >> HA	2	R2
	mass transport >> activity	2	R3
	termination eventB >> event	5	R4
	R2+R3	4	-
	R2+R4	7	-
	R3+R4	12	-
	R2+R3+R4	18	-
Nihon-zaka tunnel	remove all properties	7	-
	artifact actor >> actor	2	R1
	operation >> HA; chemical reaction >> activity	1	R2
	artificial PO >> PO; destruction event >> event	2	R3
	R1+R2	3	-
	R1+R2+R3	21	-
Leakage at Mihama	remove all properties	14	-
	delete "elemental material"	1	R1
	fluid object >> PO	2	R2
	reaction phenomenon >> activity	1	R3
	R1+R2	3	-
	R1+R2+R3	7	-
	R1+R2+R3+ machine artifact >> PO	11	-

No clear pattern emerged from replacing single classes in the base queries with corresponding upper classes. In the "Seongsu Bridge" and the "Nihon-zaka tunnel" queries, replacing the termination or destruction event with "event" had the greatest effect, while in the "Leakage at Mihama" query, only replacing specific physical objects with "physical object" resulted in increases of matches.

In almost all cases, replacing combinations of specific classes with their corresponding upper classes resulted in far more matches than the sum of the additional matches when individual classes were replaced. For example, in the "Nihon-zaka tunnel" query, replacing three classes simultaneously versus replacing them separately increased additional matches ten-fold (1+0+1 versus 20). In other words, there were 18 statements that matched when all three mid-level classes were replaced but do not match when any one of the mid-level classes was replaced. This non-linearity of query selectivity in the EKOSS system contrasts with the linearity of Boolean searches where queries with multiple terms connected by "OR" simply give the union of the matches for each individual term.

EKOSS allows a user to express the semantics of a search condition both by asserting properties and by selecting the class specificity. As examples of increased search recall, we can consider the six additional matches to the relaxed query for "Seongsu Bridge" labeled "R24" to be likely candidates for cases that did not match the original query exactly but that have similar failure mechanisms. The query "Seongsu Bridge – R24" can be expressed as:

> Some **activity or phenomenon** with a *specified* **human failure type** *has an end* **event** that is the *start of* a **mass transport phenomenon**, which *ends in* a **termination event** with a *participating* **physical object**.

One additional match is the case "Steam Eruption from Nuclear Power Plant Cooling System." The matching part can be expressed as:

> a **measuring activity** with **disregard of procedure** ended in an **artifact destruction event** causing a **leaking material transport**, which ended in a **death event** of "5 workers".

Another match is the case "Brittle Fracture of Hydrodesulfurization Reactor during Pressure Test," whose matching description can be expressed as:

> an **operation** with **insufficient understanding** caused a **multibody movement event** that caused a **mass body explosive movement**, ending in an **artifact destruction event** of the "facilities of the factory".

Neither of these cases matched with the "Seongsu Bridge" query when all properties were removed, so keyword matching would not have produced these matches. However, it is clear that they share common semantics in the mechanism by which the failure occurred. This demonstrates the benefits of the semantic search in discovering cases from other domains where the knowledge of the "Seongsu Bridge" failure could have been useful.

Several information systems for reusing knowledge about failures have been reported in the literature (Goble and Bukowski, 2007; Jacobo et al., 2007; James, 2005; Moon et al., 1998; Stone et al., 2005; Warren Liao et al., 1999). However, none use semantic inference to match failure cases with information needs. For example, although the ES-FAME system uses case based and rule based reasoning, it does not support logical reasoning to infer implied relationships (Jacobo et al., 2007). As discussed in the introduction, even existing systems for semantic matching do not support reasoning about chains of instance relationships or instances of the same class having different attributes, which we believe is necessary for effective reuse of knowledge about complex issues such as failure mechanisms.

The SCINTENG ontology was able to express the failure mechanisms in the 212 cases we used. The activity event model of the ECM gave us a way to logically extend the original case scenarios. Because failure mechanisms are classified under "class of individual", which is disjoint with classes such as "event" and "activity", the knowledge of the activity-event chain is orthogonal from the knowledge of the failure mechanisms. This allows us to search for specific failure mechanisms, specific activity-event chains, or combinations of the two. Finally, the SCINTENG ontology lets us describe in detail the physical objects in a failure case, including their relationships to events and activities in the activity-event chain, compositional relationships with each other, and their material composition.

The SCINTENG ontology is unable to express some important aspects of a failure case. Several types of the concepts commonly used in case reports, e.g. job types and event scales, are not available in the ontology. Some of the sub-trees in the ontology are not rigorously defined, particularly the physical process tree and the substance tree. Adding more complex class descriptions would reduce ambiguity and enable more powerful logic inference. Also, the granularity of the ontology could be increased to give greater matching precision. However, we must consider the tradeoff between these advantages and the increased complexity arising from the additional assumptions that are incorporated in the ontology (Buckingham Shum et al., 2007). The ability to construct semantic relationships helps to overcome the relatively coarse granularity of the vocabulary. For example, although there is only one term for "disregard of procedures" in the ontology, the particular procedure that was disregarded can be specified in detail using semantic relationships.

This paper has attempted to demonstrate the feasibility and the effectiveness of applying EKOSS to the reuse of knowledge about failures in situations where the applicability of the relevant knowledge is not obvious without the use of semantic inference. The interested reader can explore the matches to the queries presented in this paper by accessing the public EKOSS server at http://www.ekoss.org.

ACKNOWLEDGEMENTS

We are grateful to Y Nakamura who created most of the semantic statements used in this paper and to B Kemper who implemented the original version of EKOSS. We also thank Y Hatamura, H Kobayashi, M Kunishima, M Nakao, and M Tamura for advice concerning the analysis of the cases in the failure DB. Funding for this research was provided by the Knowledge Failure Database project at the Japan Science and Technology Agency and the Office of the President of the University of Tokyo.

REFERENCES

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

Blake, C., Rendall, M., 2006. Scientific discovery: a view from the trenches. In *Proc.9th Intl Conf on Discovery Science*, Barcelona, Spain.

Brown, S., 2007. Forensic engineering: Reduction of risk and improving technology (for all things great and small). *Eng Failure Anal*, 14(6): 1019-1037.

Buckingham Shum, S., Uren, V., Li, G., Sereno, B., Mancini, C., 2007. Modeling naturalistic argument- ation in research literatures: Representation and interaction design issues. *Intl J Intel Sys* 22(1): 17-47.

Ceol, A., Chatr-Aryamontri, A., Licata, L., Cesareni, G., 2008. Linking entries in protein interaction database to structured text: The FEBS Letters experiment. *FEBS letters* 582(8): 1171-1177.

Darlington, J. F., Booker, J. D., 2006. Development of a design technique for the identification of fatigue initiating features. *Eng Failure Anal*, 13(7): 1134-1152.

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: 269-307.

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

Goble, W. M., Bukowski, J. V., 2007. Development of a mechanical component failure database. *Reliability and Maintainability Symposium*, pp: 451-455.

Halaschek-Wiener, C., Kolovski, V., 2008. Syndication on the Web using a description logic approach. *J Web Semantics,* 6: 171-190.

Hatamura, Y., Nakao, M., Iino, K., 2003a. Prototyping a failure knowledge database. *Joho Shori* 44(7): 733-739 (Japanese).

Hatamura, Y., Iino, K., Tsuchiya, K., Hamaguchi, T., 2003b. Structure of failure knowledge database and case expression. *CIRP Annals – Manufacturing Tech* 53(1): 97-100.

Jacobo, V. H., Ortiz, A., Cerrud, Y., Schouwenaars, R., 2007. Hybrid expert system for the failure analysis of mechanical elements. *Eng Failure Anal* 14(8): 1435-1443.

James, M. N., 2005. Design, manufacture and materials; their interaction and role in engineering failures. *Engineering Failure Analysis* 12(5): 662-678.

JST, 2010. “Failure knowledge database” Available at: http://shippai.jst.go.jp/en (accessed 1 July 2010).

Kraines, S. B., Guo, W., Kemper, B. E., Nakamura, Y., 2006a. EKOSS: a knowledge user centered approach to knowledge sharing, discovery, and integration on the Semantic Web. In *Proc. of the 5th Intl Semantic Web Conf*, Athens, GA, USA, November 5-9, 2006.

Kraines, S. B., Batres, R., Koyama, M., Wallace, D. R., Komiyama, H., 2006b. Internet-based integrated environmental assessment: using ontologies to share computational models. *J. Industrial Ecology* 9: 31-50.

Kraines, S.B., Koyama, M., Batres, R., 2006c. An emergent simulation modeling approach for discovery of knowledge on phenomena in chemical systems. *J. Chemical Engineering of Japan*, 39: 1010-1027.

Li, L. Horrocks, I., 2003. A software framework for matchmaking based on semantic web technology. In *Proc. of the 12th Intl Conf on World Wide Web,* Budapest, Hungary, pp: 331-339.

Marcondes, C. H., 2004. From scientific communication to public knowledge: the scientific article Web published as a knowledge base. In *M. Dobreva, J. Engelen (Eds.), Proc 8th ICC/IFIP Intl Conf on Electronic Publishing*, Leuven, Belgium, pp:119-126.

Moon, Y. B., Divers, C. K., Kim, H.-J., 1998. AEWS: an integrated knowledge-based system with neural networks for reliability prediction. *Computers in Industry,* 35(2): 101-108

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

Stone, R. B., Tumer, L. Y., Wie, M. V.,2005. The function-failure design method. *J Mech Design* 127(3): 397-407.

Tamura, M., 2003. Learn from failure! failure knowledge in chemical substances and plants and its use. *Chemistry* 58(8): 24-29 (Japanese).

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

Wang, J., Jin, B., Li, J., 2004. An ontology-based publish/subscribe system. In *H.-A. Jacobsen (Ed.): Middleware 2004, LNCS* 3231, pp: 232-253.

Warren Liao, T., Zhan, Z-H., Mount, C. R., 1999. An integrated database and expert system for failure mechanism identification: Part I: automated know- ledge acquisition. *Eng Failure Anal* 6(6): 387-406.

Wintle, J. B., Pargeter, R. J., 2005. Technical failure investigation of welded structures (or how to get the most out of failures). *Eng Failure Anal* 12(6): 1027-1037.