Semantic Web Application in Requirement Tracing for Product Design Information

Tanasaran Trichachawanwong, Ao Gao, Xili Chen, Tomohiro Murata

Abstract— **This paper proposes the application of Semantic Web (SW) technique in the Requirement Traceability (RT). The idea begins with information flow model of RT-related document in a firm and the meta data which represents model of document information transformation is applied. Finally, the proposed RT model is implemented to experimental system by using semantic Web tool for feasibility test with an RT scenario. Query Language is used for finding the requirements to be traced and its related information inside the document data storage.**

Index Terms—**Requirement Tracing, Information Retrieval, Document Data Storage, RDF.**

I. Introduction

In the growing of business competitors, success in the business requires the full management of firm's resource to produce the most suitable product in the most affordable design and manufacturing within the limited time. To achieve this goal, Information system for Requirement traceability (RT) attracts attentions. This paper proposes a Semantic Web application as an information system tools for RT system. Prior to this paper, some RT system was proposed by studies such as: [4] that proposes a system taking priority on "Trace Value," or [13] that tags paragraphs related to RT. Presently, to manage change of product design information, firm use PDM (Product Data Management) tools which manage the product design information with version management. From now on, emphasizing more on improving customer satisfaction, requirement traceability is needed for a firm to grasp every product design changes distributed in each responsible firm unit to adapt their manufacturing activities properly and regularly to the user requirement change.

II. Requirement Traceability

RT is the ability to follow the life of a requirement from its origins through its development and specification [6]. Some definition mentions that RT is a link or definable relationship between entities [8] and some states that it is the links of realities and the models between people and system [5]. From perspective of production information system for RT, it is more understandable definition that RT is the ability to look for related information through an information system of meaningful links that related each piece of traceable information of product requirement and manufacturing together.

RT process can be divided into 2 domains at "Requirement information" shown as the square of Requirement in the diagram of Fig.1. Left side domain of the square of Requirement is called as Backward tracing zone where the information emerged before the requirement is traced, such as the origin of the requirement, the variation of the demand, or the thought of the stakeholders. Right side of the square of Requirement is called as Backward tracing zone where the information emerged after the requirement is traced, such as information of design and production. In this paper, we focus on Forward tracing because it is important and effective to trace information transformation process from requirement, information such as documents that is related to the design, and production/implementation to improve the human cooperation in the design process, such as the cooperation of the stakeholders or customers in the implementing of the product design and in the manufacturing process in the firm.

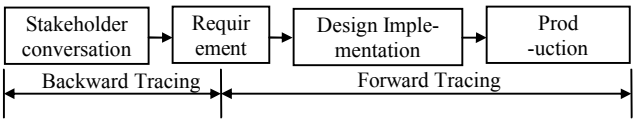

Fig. 1 Backward and Forward Tracing

The objectives of RT can be summarized based on what suggested by [5] below:

- Stakeholder Goals or contribution to the development and satisfaction of requirement
- Requirement and its linkage to concerning topics.
- Design Implement
- Design Complement Information or design rationale composes of decision made, alternative considered, and underlying assumptions
- Manufacturing information
- Manufacturing Complement Information such as task performed, plans followed and resources consumed

Usage of the RT system, which emphasizes more on connecting each pieces of key information together, hence, increase the value of the information inside the firm information system. It also helps uncover unexpected problems and innovative opportunities, and lays the groundwork for corporate knowledge management [5]. On

Manuscript received December 30, 2007.

Tanasaran Trichachawanwong, Ao Gao and Xili Chen are master students in Graduate School of Information Production and System, Waseda University (2-7 Hibikino, Wakamatsu-ku, Kitakyushu-shi, Fukuoka, Japan).

Tomohiro Murata is a professor in Graduate School of Information, Production and System, Waseda University, Japan (2-7 Hibikino, Wakamatsu-ku, Kitakyushu-shi, Fukuoka, Japan.

the other hand, each pieces of these key information are usually located in several data resources with heterogeneously distributed manner. Semantic Web is a promising technique to link pieces of key information effectively and flexibly.

III. Semantic Web-Based RT System

A. Architecture of Semantic-Web- based RT system

One of the SW objectives is to enhance the ability of both people and software agents to find documents, information and answers to queries on the web [12]. RT system is also said to be part of the Information Retrieval problem [2]. Fig.2 shows architecture of RT system based on SW. Storage & Inference layer (SAIL) stores meta data for RT based on RDF model and infer chains of linkage among requirement information, design information and manufacturing information through the usage of RDF triple [7], which composes of: Subject, Predicate and Object. Users from each division access the system; add, delete and query RT information.

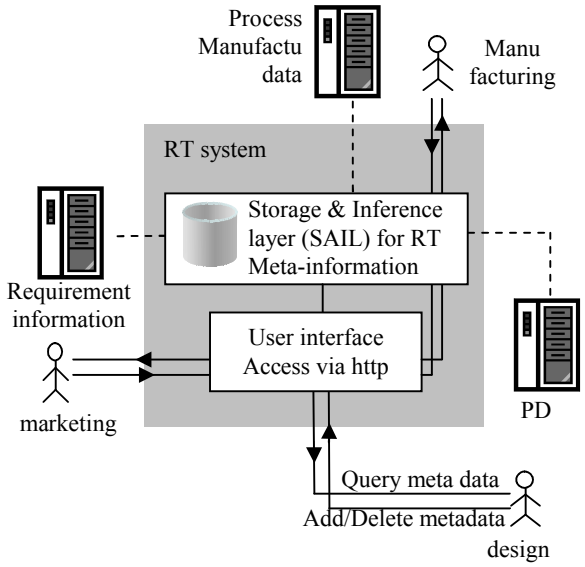

Fig. 2 Architecture of RT system

B. Information Model for RT

Applying semantic web data storage and developing an RT system, there are topics, required to be discussed until the RT system is materialized:

- Information Flow model of RT
- Design Rationale Information
- Tag of Information Linkage
- Classification of document type
- Relation model of Product Information

1) Information Flow model of RT

Fig. 3 illustrates general information flow of RT. In the information flow, two types of document are kept in the firm information system. The First is document already stored in each database—ready to be accessed. The Latter is that not being kept systematically in the information system and might not be traceable. Requirement, Design Rationale and Process Management Information can be categorized into the latter group. They, at least in this thesis, need traceability. So, this paper proposes that, the unmanaged information—written in balloon shape, to be materialized into formal database (written in cylinder—as of CRM, Specification, PDM and Process Manufacturing Data) that formally keeps information about them, so that firm personnel can access them, through links.

Information from the customers flows from marketing, design into Manufacturing Division. Marketing arranged the product-change or new product talks with the stakeholders (or customers). This talk is formally documented and submitted to CRM (Customer Relationship Management) system which is dedicated to store that of customer. When stakeholder updates his requirement, the information keeps upgrading and new document object is produced and stored into the CRM server.

The analyzed requirement, each has links to its related "sub" requirement. When manager decides that some requirement needs to be adjusted, changes take place in the requirement [10] and the new requirement is created based on the old one. Links between the old requirement and the new one are recorded. Specification is drafted to guide the designers the summary of all requirements needed to be satisfied. It gives brief descriptive information on the parts or component attributes requirement. Specification provides chances to marketing manager whether any requirement has been unallocated.

The responsibility on the RT system of marketing division ended with draft specification. To be transferred to next information flow, specification is used in the design implement under design division. Designer can have some inquiry about the specification, "Why the specification is written like this?" In this case, RT system is needed. Modeled in Fig. 3, design division composes of 2 servers which are PDM (Product Data Management) and Design Rationale Information server. Designer has to trace the requirement that links with the given draft specification to find out the answer. Designer takes specification into main consideration when implements design. Each design document is stored in PDM.

When design document undergoes revising or reorganizing, the obsolete one should still be kept reference. The later version is linked with the previous one, in that tracking from the later version to the previous version can be possible and gives designer some general idea on that component design history.

2) Design Rationale Information [5]

Information such as a decision made during design, alternatives considered, underlying assumptions are included into RT system.

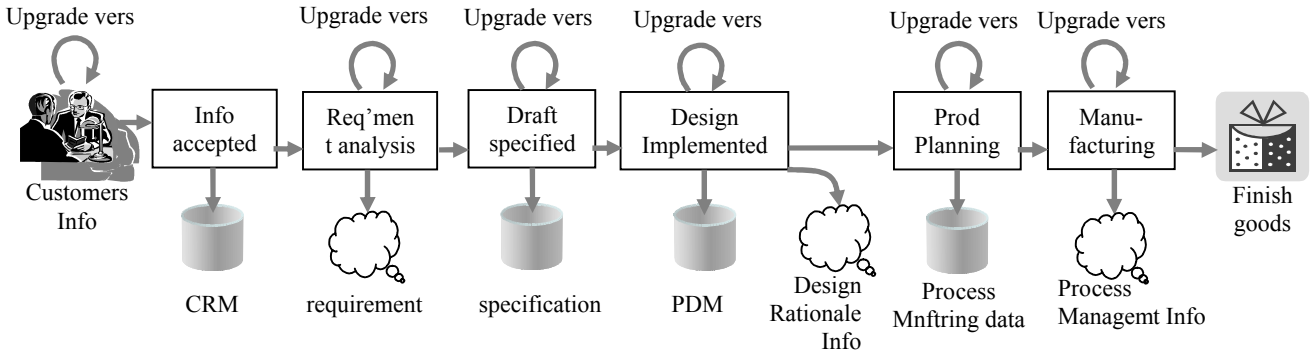

Fig. 3 RT-related information flow model

Design Rationale Information is proposed to store in another server located inside the design division, accessible to traceability. Then, design information flows to manufacturing division which arranged the material and schedule for production control.

Production manager prepare the data for manufacturing. This data is kept in the process manufacturing data. Beside this document, there is Process Management Information which is included into RT system. Process Management Information keep traceability document such as, tasks performed, plans followed, and resources consumed [5] during the specific production line.

3) Tag of information linkage

The unmanaged information, i.e. requirement, Design Rationale Information and Process Management Information, proposed to be kept in each database, represented in cylinder shape. A tag, in this research, means a meaningful word uses to describe the relationship or links between 2 objects. A tag can also describe the relationship between an object and its meta data, piece of information that give information about other information. Document object is a piece document discuss about a single topic, stored in any accessible (traceable) database. Object can be file stored in the database, regardless of its file types, size or etc.

4) Classification of document type

Table 1 shows classification of document type for RT. Any document can subject to updating, revising and obsolescing.

Table 1 RT-related document classification

Division	Doc type	Document Name
Marketing	CR	Stakeholder Goal
	MR	Requirement
	SP	Draft Specification
Design	D	Design Document
	DC1	Decision Made During Design
	DC2	Alternative Considered
	DC3	Underlying Assumption
Manufacturing	M	Process Manufacturing Data
	MC1	Tasks Performed
	MC2	Plans Followed
	MC3	Resources Consumed

5) Relation Model of Product Information

11 document types of product information and 12 relation properties are proposed to represent inside RT system linked document instances. Fig. 4 shows the model of information transformation of the proposed product information and types of document object are shown as an oval.

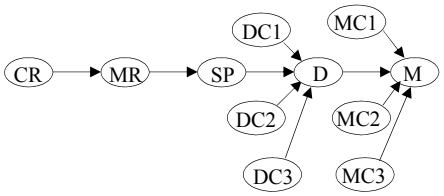

Fig. 4 Model of Information Transformation of Product Information

Typically, one oval means one piece object of document or file. A document has relations with others. The relation tags have distinct names. However, one oval also means several pieces of object, for the case of MR; one oval of MR can mean a branch of many MR, as requirement is usually in large amount of information or files. Meta tag is written in arrow with meaningful relation words attached. There are two kinds of meta tag of "Document to Document Meta tag" and "Document to Literal Meta tag" and these are summarized in Table 2.

Table 2 Proposed Relation and its Connection

REL TYPE	DOC	RELATION	DOC
Document to Document relation	CR	requires	MR
	MR	specifies	SP
	SP	constrains	D
	DC1	isDecidedFor	D
	DC2	isChosedFor	D
	DC3	isAssumedFor	D
	D	isMaterializedInto	M
	MC1	isPerformedFor	M
	MC2	isPlannedFor	M
	MC3	isConsumedFor	M
	a doc	isReferencedBy	another doc
	a doc	hasVersion	another doc
Document to Literal relation	a doc	"tag"	"Literal"

For example, "requires" tag is used to link between (CR to MR or MR to MR) One requirement links to another requirement using "requires" tag, making the requirement, naturally large amount, connect with each other in branch-like pattern. CR is located in the branch as the stem of all requirements. "hasVersion" tag (a document another

document) this tag explains the relation that one document on the arrow head is the new version of the other on the arrow tail. Two document objects should be same type.

"tag" of "Document to Literal Meta tag" (a document to a literal Meta tag) is used to explain the document on the arrow tail, what keyword, explanation word, remember word, etc. One document can have multiple "tag".

One document object need to discuss only 1 topic, such as MR1.doc discusses about requirement to decrease the Sensitivity of microphone. When a document become defunct (its new version came out making the old one defunct), its relation to other document should be shift to the new version of the document too.

6) Example of Relation Model of Product Information and RDF description

Fig. 5 shows an example scenario the basic structure of which is compliant with the relation given in Fig. 4 and each pair of document object is given appropriate link as stated in Table 2. The RT scenario of product design information is given as follows. The scenario starts from a record of customer conversation, namely CR1.doc, Marketing staff then recorded this requirement into CR1.doc (a Word document) and store the information accessible to relevant staffs. The requirement emerges, as it is needed to adjust the microphone sensitivity—MR1.doc. Some other personnel analyzed the situation and proposed that there are 2 ways that has to be carried out together, in order to fulfill requirement stated in MR1.doc. First, MR100.doc discusses that the noise reducer circuit is required to be integrated into the phone. Second, MR150.doc requires that the microphone sensitivity should be reduced. Coming to this step, we can show that pairs of requirement have their relationship. MR100.doc and MR150.doc are both emerged from MR1.doc. Meanwhile, MR1.doc requires both 2 requirement to be accomplished so that MR1.doc can be satisfied. The relation of CR1, MR1, MR100 and MR150 are shown in Fig. 5. Its associated RDF graph is shown in Table 3, written partially in RDF/XML. Document to Literal Meta data, rqt:tag, are shown in Table 3 such as "high surrounding noise" or "microphone sensitivity".

Table 3 CR1, MR1, MR100 and MR150.doc relation expression in RDF/XML

```
  <rqt:CR
rdf:about="http://133.9.41.150/mkt/cr/CR1.doc">
  <rqt:requires rdf:resource=
"http://133.9.41.150/mkt/mr/MR1.doc"/>
  <rqt:tag>complaint</rqt:tag>
  <rqt:tag>microphone sensitivity </rqt:tag>
  <rqt:tag>surrounding noise</rqt:tag>
  </rqt:CR>
  <rqt:MR
rdf:about="http://133.9.41.150/mkt/mr/MR1.doc">
  <rqt:tag>high surrounding noise</rqt:tag>
  <rqt:tag>microphone sensitivity</rqt:tag>
  <rqt:tag>edit the sensitivity</rqt:tag>
  <rqt:tag>edit</rqt:tag>
  <rqt:requires rdf:resource=
"http://133.9.41.150/mkt/mr/MR150.doc"/>
  <rqt:requires rdf:resource=
"http://133.9.41.150/mkt/mr/MR100.doc"/>
  </rqt:MR>
```

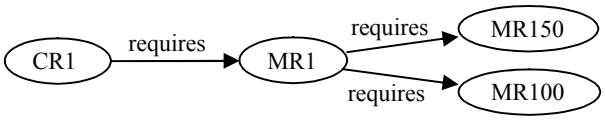

Fig. 5 Sample of Connection as of Table 3

IV. An Example and Experiment

A. Example Scenario for feasibility study of the proposed semantic-web-based RT system.

Fig. 6 gives a whole scenario of branch of document relation ramifying from CR1.doc, which includes the example scenario in Fig.5 as a sub scenario, will be used for sample for feasibility study of RT query execution shown later in this chapter. Brief description of each object of document or file in Fig. 6 is given in Table 4. The RT query is "The stakeholder requested to a mobile phone company, that the designed phone model has a problem. User's voice cannot reach the mobile phone well where the environment is noisy"

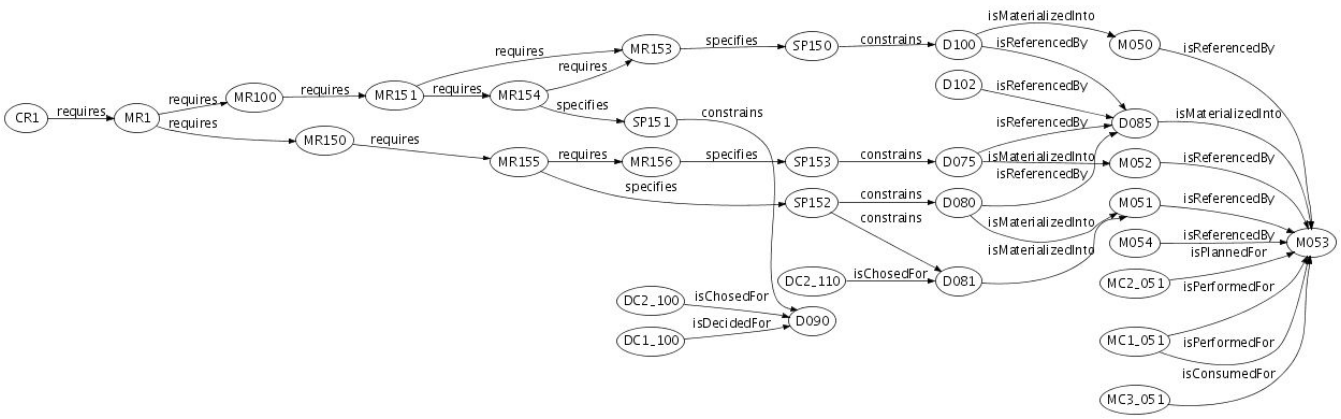

Fig. 6 View of each connection used in RT scenario of this paper

Table 4 Document list and its description

FILE NAME	BRIEF DESCRIPTION
CR1	January 10 2007, Customer discussed that, the sensitivity of the microphone is inappropriate. Third person voice or surrounding noise seems to enter the microphone, causing user to speak louder or cover the microphone to keep the noise down.
D075	Phone plastic body
D080	Phone Microphone sound absorber portion detail
D081	Sound absorber engineering detail
D085	Assembling detail
D090	Electric Circuit with noise reducer
D100	Printed Circuit Board
D102	Phone Button Design
DC1_100	Decided to assign chipset design for noise reducer to subcontract
DC2_100	in 4 different chipset companies', company X was chose because (more)
DC2_110	In 5 manufacturers, we choose material from X because (more)
M050	Production of PCB
M051	Production of noise reducer
M052	Production of plastic body
M053	Production detail of assembling
M054	Production of others
MC1_051	Performing the noise reducer test both on finishing phone and raw material
MC2_051	Using production plan as given (more)
MC3_051	Opening Saturday shift to shorten the cycle
MR1	To edit the sensitivity of the microphone, in case the surrounding noise is high
MR100	to add noise reducer circuit
MR150	to reduce the microphone sensitivity
MR151	to edit microphone part by adding noise reduction
MR153	to redesign of PCB
MR154	to change to voice processing chipset with noise reduction
MR155	to add voice reducer material
MR156	to redesign plastic body with voice reducer portion
SP150	specification of PCB
SP151	specification of noise reduction circuit
SP152	specification of noise reduction material
SP153	specification of plastic body

The responsibility of inputting these relations into the Sesame Repository Server bounds to each personnel that takes responsibility for each type of RT-related document. Each personnel accesses the RT system, inputs the relation, edit or update the relation, as commented by [5].

Relation of each document continues and ramifies into a branch of related document. Their file lists and description are given in Table 4.

B. Sesame Framework

This study chooses Sesame (Sesame-Memory so to say) as a SW server tool not because its highest performance of [1], but open-sourced nature, and web forum Q&A. Sesame, rdf framework that supports RDF Schema inferring and querying, can be deployed as a Java Servlet Application in a web server that supports Java Servlet and JSP such as Apache Tomcat. Users add, delete and query data that executed in "Repository", a storage container for RDF, which we called it "Storage and Inference Layer for RT Meta Information".

The 7 capability RT tools [2] should be met: (a) identify each requirement, (b) assign a unique identifier to each requirement, (c) allocate all children requirements present in the lower level document, (d) allocate parent requirement in the high level document, (e) examine each high level traced requirement and determining if it has been completely satisfied by the low level requirements that were selected as links, (f) prepare a report that presents the traceability matrix, (g) prepare the report on percentage of the high level requirements were completely satisfied. These criteria are used as a guideline to the RT scenario below. Using SW technique, this paper proposed RT system that can meet topics (a) through (e).

URI (Uniform Resource Identifier) of a resource is its identifier. Because, 2 identical file names cannot be placed in the same directory, hence, this system need not generate any ID—using its URI as an existed one.

Children requirement is another requirement that is downstream to a specific requirement. Due to the number and hierarchical structure of requirement document, it is rather not easy to say how many hierarchy each requirement document is in, we proposed that, the link used to link each requirement, be in transitive property, so that, the inference engine can infer a new relation, causing the tracing for all children requirement to be carried out in a single query.

C. Requirement Tracing query results for the scenario

Using SeRQL query language for Sesame [11] the query will look similar to Table 5.

Table 5 Children Requirement of MR100.doc

```
SELECT child FROM
{http://133.9.41.150/mkt/mr/MR100.doc} rqt:requires
{child}
```

From Table 5, MR151, MR153 and MR154.doc will be traced out.

For each low level requirement, locate a parent requirement (upstream) in the high level document. Examine each high level traced requirement and determine if it has been completely satisfied by the low level requirements that were selected as links. Table 6 shows how to query.

Table 6 Parent Requirement of MR154.doc

```
SELECT Parent FROM
{Parent} rqt:requires
{http://133.9.41.150/mkt/mr/MR154.doc}
```

CR1, MR1, MR100, MR151.doc will be answered as upstream document.

The other similar tracing problem to Children/Parent Requirement is, "to find all document downstream of the

workflow that has impact, when MR152.doc is changed or canceled. As discussed in our proposed information model, the requirement (MR) is connected with Specification (SP). SP is connected with D. And, D is connected with M respectively. The query for downstream of the information flow is shown in table 7.

Table 7 All downstream document of MR155.doc

```
   SELECT r,refr,s,refs,d,refd,m,refm FROM
   {http://133.9.41.150/mkt/mr/MR155.doc}
rqt:requires {r} rqt:specifies {s} rqt:constrains {d}
rqt:isMaterializedInto {m},
   [{r} rqt:isReferencedBy {refr}],
   [{s} rqt:isReferencedBy {refs}],
   [{d} rqt:isReferencedBy {refd}],
   [{m} rqt:isReferencedBy {refm}]
   UNION
   SELECT r,refr,s,refs,d,refd,m,refm FROM
   {http://133.9.41.150/mkt/mr/MR155.doc}
rqt:specifies {s} rqt:constrains {d}
rqt:isMaterializedInto {m},
   [{http://133.9.41.150/mkt/mr/MR155.doc}
rqt:requires {r}],
   [{r} rqt:isReferencedBy {refr}],
   [{s} rqt:isReferencedBy {refs}],
   [{d} rqt:isReferencedBy {refd}],
   [{m} rqt:isReferencedBy {refm}]
```

After querying for the RT with Sesame based experimental RT system, the web browser screen will display the result of RT as of Fig. 7 in table format (Sesame can also display in RDF or XML format as well). The tracing result will be: MR156, SP152, SP153, D075, D080, D081, D085, M051, M052 and M053.doc, which is an expected result.

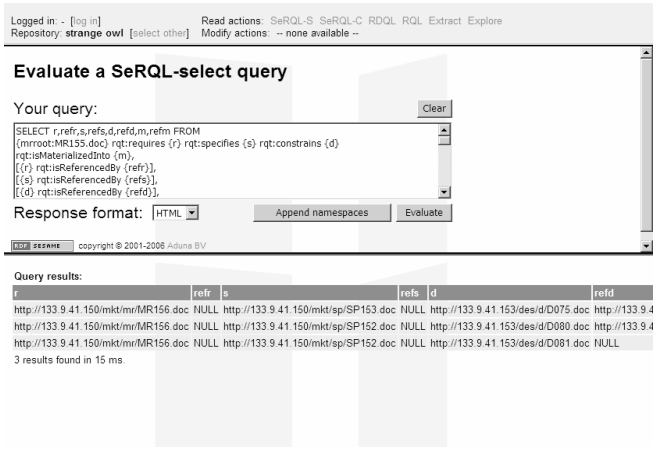

Fig. 7 Query result from Table 6 in a Web Browser

V. Pros and Cons

This RT system applies SW data storage as a web service RT system. On the concept [3] that, RT tool should take care of time-consuming parts of indexing each document's pair relation. Proposed system could then allow the analyst to concentrate more on the parts that really requires human-decision. Emphasizing on SW data storage technology, this tool can retrieve document that hidden in complicated relation with other document. Sesame framework, once installed into a SW data storage server, is accessible on intranet; hence, it has no cost for installation into each network computer, nor OS conflict.

The weak point of this system is that, all users need to understand the basic of writing RDF/XML as for inputting the relation to the Repository, and the Sesame Query Language (SeRQL) as for querying RT. Also, this tool still lacks convenient user interface.

References

[1] B. Liu, B. Hu, "An Evaluation of RDF Storage Systems for Large Data Applications," skg, p. 59, First International Conference on Semantics, Knowledge and Grid (SKG'05), 2005

[2] J. H. Hayes , A. Dekhtyar , S. K. Sundaram , S. Howard, "Helping Analysts Trace Requirements: An Objective Look," Proceedings of the Requirements Engineering Conference, 12th IEEE International (RE'04), p.249-259, September 06-10, 2004

[3] J. H. Hayes, A. Dekhtyar, S. K. Sundaram, "Advancing candidate link generation for requirements tracing: the study of methods," Transactions on Software Engineering, vol.32, no.1, pp. 4-19, Jan. 2006

[4] M. Heindl , S. Biffl, "A case study on value-based requirements tracing," Proceedings of the 10th European software engineering conference held jointly with 13th ACM SIGSOFT international symposium on Foundations of software engineering, September 05-09, 2005, Lisbon, Portugal

[5] M. Jarke, "Requirements tracing," Communications of the ACM, v.41 n.12, p.32-36, Dec. 1998

[6] O. C. Z. Gotel, A. C. W. Finkelstein, "An Analysis of the Requirements Traceability Problem," Proceedings of International Conference on Requirements Engineering 1994, IEEE CS Press, (1994), 94-101.

[7] O. Lassila, R. Swick, (1999, January 5). Resource Description Framework (RDF) Model and Syntax Specification [Online]. Available: http://www.w3.org/TR/PR-rdf-syntax/#model

[8] R. Watkins and M. Neal, Abbott Laboratories Diagnostic Division, "Why and How of Requirements Tracing," IEEE Software, vol. 11, no. 4, pp. 104-106, Jul/Aug, 1994.

[9] Shelley Powers, *Practical RDF*. CA: O'Reilly & Associates, Inc., 2003, ch. 10.

[10] S. D. P. Harker, K. D. Eason, and J. E. Dobson, "The change and evolution of requirements as a challenge to the practice of software engineering," Proceedings of the IEEE International Symposium on Requirements Engineering, Los Alamitos, California: IEEE Computer Society Press, January 1993, pp. 266-272.

[11] The Sesame and OMM development teams. (2007 December 12). *Chapter 6. The SeRQL query language (revision 1.2)* [Online], Available: http://www.openrdf.org/doc/sesame/users/ch06.html

[12] T. Finin, J. Mayfield , A. Joshi , R. S. Cost , C. Fink, "Information Retrieval and the Semantic Web," Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 4, p.113.1, January 03-06, 2005

[13] X. Song, B. Hasling, G. Mangla, B. Sherman, "Lessons Learned from Building a Web-Based Requirements Tracing System," Third International Conference on Requirements Engineering (ICRE 1998), p. 41-50