EXPLOITING NEIGHBORHOOD AND MULTI-DIMENSION GRANULAR INFORMATION FOR SUPPORTING DESIGN RATIONALE RETRIEVAL

Yan Liang[1], Ying Liu[2], Wing Bun Lee[1], Chun Kit Kwong[1]
(1) Hong Kong Polytechnic University, HK (2) National University of Singapore, Singapore

ABSTRACT
Based on our previously proposed ISAL model (issue, solution and artifact layer) for design rationale (DR) representation, in this paper, we report our efforts in researching an ISAL based DR retrieval framework to better support DR retrieval by taking advantage of neighborhood and multi-dimension granular information presented in DRs. In our proposal, DR is firstly extracted and indexed using ISAL and a document profile model respectively. Next, an initial DR graph is formed by linking up different DRs based on their document citations and document similarities. A DR network is therefore established by integrating similarities from issues, solutions and artifact aspects using neighborhood information in the DR graph. In order to prioritize DRs retrieved, a graph-based ranking approach is further engaged. To validate the approach proposed, we have reported our preliminary experiments on issues like DR indexing based on different approaches, similarity measurement in DR network, and lastly, a brief example of using neighborhood information to suggest potential DR related concepts in retrieval query processing.

Keywords: Design rationale retrieval, Design rationale network, Neighborhood information, Granular information

1 INTRODUCTION
The rapid development and use of information technology have made a large amount of design information and knowledge digitalized. Digital design documents can be obtained from both external sources, e.g. patents, journal and magazines, and internal sources, e.g. design reports, drawing notes and design logbooks. They are sources of critical design information that has received enormous interests and studies in engineering design, such as technology opportunity analysis and trend analysis. With the ever-increasing number of design archival documents, it has become challenges for organizations and designers to retrieve useful design information effectively for design reuse, innovation and creativity.

One stream of such critical design knowledge is design rationale (DR) that, generally speaking, explains why an artifact is designed the way it is [1]. It captures relevant issues to understand design objectives, records design assumptions and constraints, and helps analyze reasons and arguments behind the artifacts. If DRs of previous designs can be stored and be effectively accessible, it can help designers to understand the design know-how and technology, and better utilize design knowledge to fulfill specific design purposes.

Many DR systems have been developed for engineering designers to capture and record DR information, such as SEURAT for software development [2], DRed [3] and AREL for software architecture traceability [4]. However, existing DR approaches require heavy human involvement that is time consuming and labor-intensive. Moreover, they can only start to record the rationale by designers along the design processes. For those DRs that are stored in design archival documents such as design reports and patent documents, it requires much effort to transfer them into a DR system. In addition, the manual DR process manner cannot afford to timely handle the DRs stored in ongoing design archival documents.

The research on DR retrieval receives less attention compared with studies on DR capture [5]. Existing DR systems use formal graphs to represent DR. In DR graphs, nodes often represent rationale elements, such as issues, problems, positions, arguments and constraints, and links between nodes indicate their relationships. In DR retrieval, one basic approach is navigation, which allows designers to explore DR and investigate the details by traversing from one node to another by existing links [1].

However, this navigation process becomes complicated especially when the size of DR graphs is relatively large. Another basic retrieval approach is to provide query-based search, which aims to locate relevant DR pieces in response to a given query [1]. It is relatively efficient than browsing the nodes of DR graphs. However, the exact matching of query keywords may be insufficient to rank all the relevant information, since some information needs are quite fuzzy and the rationale relationship is often neglected in the retrieval process. In addition, it would be better to provide supports for result navigation when the retrieved results are quite large.

Based on the observation of existing DR approaches, we have proposed a computational DR representation model ISAL (issue, solution and artifact layer) that aims to discover and restore rationale information from design archival documents particularly with significant textual content [6]. In this paper, we focus on DR search and retrieval from a large amount of design archival documents based on our ISAL model. We propose a retrieval strategy using neighborhood information in the DR network and multi-dimension granular information to better support DR information retrieval. The rest of this paper is organized as follows. Section 2 presents the related work of DR representation, DR retrieval and design document retrieval. In Section 3, we introduce the overview of our DR retrieval strategy. In Section 4, we detail our methodology of the DR retrieval by exploiting neighborhood and multi-dimension granular information. Section 5 shows the experiments on indexing and DR network construction. Section 6 concludes the paper.

2 RELATED WORK

In DR studies, how to represent and model DR is a fundamental issue to establish DR capture systems as well as DR retrieval strategy. The argumentation-based approach is the mainstream approach to represent DR. The earliest proposal specific for DR method is the Issue based Information System (IBIS) [7] and it is still the basic model for some other DR representation methods. It uses rationale elements as nodes, e.g. issue, position and argument elements, and uses predefined relations to link up rationale elements. For example, the “respond” relationship between an issue and a position indicates that a “position” is addressed to an “issue”, and the “support” relationship between a position and an argument shows that an “argument” is positively support the “position”. Several graphical DR systems have been implemented for DR capture based on some concepts of IBIS method, such as gIBIS for graphical IBIS [8], DRed [3] and Compendium tool [9]. In addition, several other DR representation models have been derived from IBIS. For example, Procedural Hierarchy of Issues (PHI) broadens the scope of “issue” and modifies the relations [10], and Kuaba extends IBIS by introducing an ontology vocabulary which defines a set of properties and relations [11].

In DR information retrieval, navigation is a common way to browse rationale information based on the nature of the DR representation model. For example, the graphical systems like DRed [3] and Compendium tool [9] allow designers to perform reasoning through traversing from one node to another by the links. Designers can obtain why a position is proposed by seeking from one position node to its dependency argument nodes. However, it would be difficult to navigate on a graph with large size. Hierarchical tree structure was also used to explore DR. One example is SEURAT (Software Engineering Using RATionale) [2], in which users can navigate the rationale by expanding or collapsing nodes on the tree structure. However, this tree-like view is weak in showing the relations between rationales in the leave nodes.

In addition, query-based retrieval is another retrieval approach that aims to locate the relevant rationale pieces in response to a given query. Kim et al. [12] presented a framework for DR retrieval by introducing semantic relations between elements in the DRed files. Given a key-word query and specified the semantic relations, natural language processing (NLP) techniques were used to measure the similarity between rationale nodes. Recently, Wang et al. introduced some methods to support DR retrieval in DRed [5]. They extended the key-word based search by suggesting potential keywords based on initial letters of keywords given by the designers and they measured the relevance of rationale pieces based on vector space model to represent rationale nodes.

While these approaches are shown to support rationale retrieval, there are some limitations in DR retrieval studies. Firstly, the existing DR systems can provide limited DR information since they require incremental human efforts to record the rationale. Particularly, it requires significant involvement to transfer DRs in a large amount of design archival documents into existing DR systems. However, designers often cannot afford to spend much time in annotating DRs in the design documents [13]. Secondly, the search method by matching the query keywords may not well retrieve the desired results without considering the nature of the DR network. Furthermore, a multi-dimension navigation strategy is needed, especially when the volume of DR repository increases. In order to make the DR process more efficient and tractable, we have proposed an ISAL model to discover DR from design archival documents [6].

Other relevant studies include engineering design documents processing although they are limited. In order to support designers in searching for useful design information, the techniques such as information extraction and information retrieval are applied in design document processing. The study of design document retrieval can be classified into several aspects based on the application purposes. The most typical purpose is to help designers to organize or retrieve design documents. McMahon et al. [14] defined a constrained-based classification and provided mapping between classification and term phrases. Yang et al. [15] introduced a thesauri-based approach for design document indexing using the vector space model and singular value decomposition techniques. Some other studies aim to help users to locate and access document content at the fragment level. Liu et al. [16] suggested organizing document fragments from hierarchical views, e.g. physical structure view, technical description view and logical content view. By using a constraint classification associated with hierarchical views, they proposed a computational framework for retrieval of document fragments.

Some other studies aim to support multi-facet search for locating engineering information. For example, Li et al. [17] built an engineering ontology (EO) to represent the established design and manufacturing knowledge from several properties such as product components, functions, material and shape features. Based on the EO, they introduced semantic-based engineering document retrieval framework. However, it requires continuous efforts to maintain the EO as the product designs evolved. In our previous study, we explored using semantically annotated product family ontology to provide multi-facet product information search and retrieval [18]. We recently proposed a methodology to build the semantically annotated multi-facet ontology in an efficient way using information extraction techniques [19].

3 AN OVERVIEW OF DR RETRIEVAL STRATEGY

In order to better support DR retrieval, we aim to design an effective retrieval strategy that enables designers to search for and navigate DR information from a large amount of design archival documents based on the ISAL model. Figure 1 shows the framework of the DR retrieval strategy.

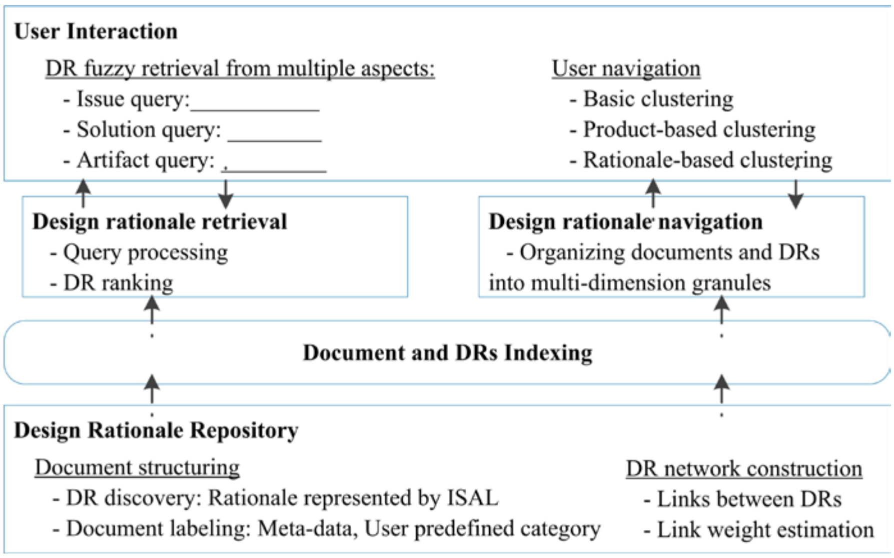

Figure 1. The Framework of DR Retrieval Strategy

The DR retrieval framework includes four important modules, namely the DR repository construction, DRs and documents indexing, DR retrieval and DR navigation supports. The DR repository module aims to organize the documents that are loaded in the system and to build up DRs and their relationships based on the ISAL model. On the one hand, in DR repository construction, document-structuring module manages to transfer each e-design document into a structural manner. It captures and discovers DR information from a document based on our ISAL rationale representation model. In addition, each design document is associated with other properties for design analysis purposes, such as metadata like date and product type, and user predefined category. Users can also define categories to classify documents in terms of their contents, such as material, manufacturing and process aspect. On the other hand, in the DR repository construction, the DR network is built by linking up the relevant DRs based on their content similarity and citation information.

The indexing module performs a process of associating or tagging documents with different search terms so that the relevant pieces of information can be located efficiently. We index the DR information from multiple dimensions. It includes the rationale aspect indexing from issue, solution and artifact layers, the product aspect indexing from product type and product structure, and the basic aspect indexing based on the metadata and user predefined categories. In the existing DR search approaches, indexed terms are often used for retrieval, but there is little consideration of relationships between DRs. In our search strategy, we intend to exploit neighborhood information of DR network to suggest potential queries and prioritize the retrieved results for retrieving relevant DR information.

Based on the document indexing and DR network, the DR retrieval and navigation modules are designed to facilitate designers' search and browsing in response to users' actions they performed. The DR retrieval module allows designers to search for DRs from several facets, i.e. issue, solution and artifact. The DR navigation module helps to guide designers in either browsing the rationale in the whole DR repository or organizing the retrieved results from multi-dimension. In this paper, we introduce our methodology of DR retrieval using neighborhood information and multi-dimension granular information.

4 METHODOLOGY

4.1 DR Representation and Indexing

The document structuring process manages to represent and interpret a design document of a collection into XML file in terms of DR information based on the ISAL model and their properties like metadata and user-predefined category. The ISAL model is a generic model for DR representation, which can support both manual DR annotation and DR discovery using text mining techniques. Our ISAL model includes three layers. They are issue layer that represents the motivations of a design, solution layer that includes the corresponding solutions, its effects and the arguments, and the artifact layer that refers to the artifact components and properties [6].

Firstly, the DR information r_i of a document d_i is represented based on our ISAL model. On the one hand, we support users to annotate the DR in a document into the ISAL structure. On the other hand, we can use text mining techniques to extract DR in the ongoing free texts into ISAL structure. For the issue layer, we have improved a graph-based ranking algorithm to extract issue-bearing sentences by using language patterns and sentence relationships based on their motivational semantics. For the solution layer, we have proposed to build up two sentence graphs for solutions and reasons respectively and use language patterns and information propagation techniques to extraction solution and reason bearing sentences. For the artifact layer, we have extended our document profile (DP) model for artifact information extraction by ranking terms based on their positional and mutual information. Using text mining techniques for DR information discovery is an alternative way to extract DR information from design archival documents in a tractable manner.

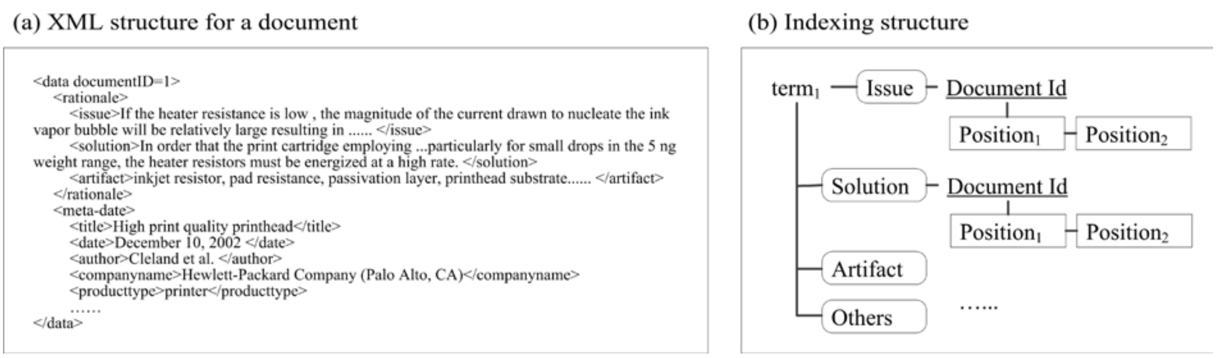

Figure 2. The Snapshot of XML Structure for a Document and its Indexing Structure

Secondly, the document labeling process is performed to tag each document with several properties like metadata and product classification for design analysis purposes. The metadata includes date, author, company's name and project title. Some external design documents, such as patents and journal articles, include a reference list that explicitly indicates the linkages with other design documents and therefore will be useful to suggest the links between DRs. Design documents can also be tagged with product classification. The examples are the product type and product structure to indicate what kinds of products or product components are the concerns of a document. Moreover, the document labeling allows designers to define categories from other concepts, such as from manufacturing and green design aspects. This document labeling process can help to attach additional information to documents for better DR retrieval. After DR extraction and document labeling, each document is stored using a XML structure as shown in Figure 2(a). A XML file includes the DR information based on our ISAL, the document labeling information and the original document.

In document indexing, it is essential to identify what terms should be used to index and represent the documents for search purposes. In the vector space model, documents are basically considered as bag-of-words (BoW), i.e. each unique word in the document is considered as the smallest unit to convey information. In our study, the indexing term set T includes both single words and frequent phrases, since we consider that phrases to some extent can provide more semantic meanings than what single words can convey. We use our DP model to generate frequent word sequences as phrases [18]. Figure 2(b) shows the data structure that is used to index terms for rationale search purposes. If a term appears in a DR, its positional information will be recorded under the corresponding segment information, i.e. "issue", "solution", "artifact" segments; otherwise, its positional information will be recorded in the "others". The positional information of a term includes the document ID, sentence ID and location of the sentence.

4.2 DR Network Construction Using Neighborhood Information

In the existing DR modeling, it is common to represent rationale by specifying the relationships between rationale elements. For example, an issue element can be connected to other two issue elements using the "specified" relation. It indicates the relevancy between those issues. In our ISAL model, the sentences in the issue layer may well contain terms or phrases that represent the correlated motivations of why the designers intend to focus on this design. For example, in inkjet printer design, the issue of "improving print quality printhead" may possibly be associated with concepts like "high drop generator density" and "high quality print output". Therefore, our DR network construction process aims to build up the linkages between DRs and measure their connections by making use of the neighborhood information. We intend to utilize the information of the DR network for the purposes of suggesting potential query expansion and ranking the relevant DR information.

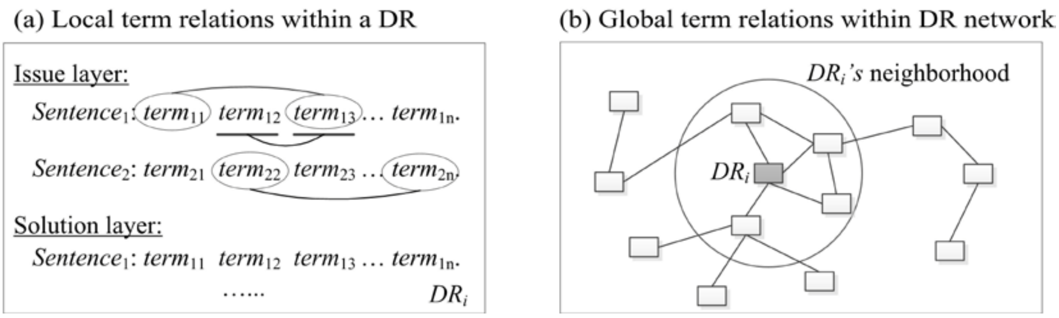

Figure 3. A Neighborhood Defined in Term Space

We define the DR network as a graph $G(R, E, \boldsymbol{W})$ to model the rationale relations of a document collection. R presents the rationale set of the collection, in which each r_i is represented as a node in G. r_i denotes a DR extracted from the document d_i based on our ISAL model. The linkage E between DRs can be built based on the reference list specified in the XML file. If there is a linkage between rationale r_i and rationale r_j, then $e(r_i, r_j) = 1$; otherwise, $e(r_i, r_j) = 0$. We can also use document similarity to link up DRs. For example, if the document similarity between d_i and d_j is larger than a threshold, than we can set $e(r_i, r_j) = 1$; otherwise, $e(r_i, r_j) = 0$. In addition, we allow users to define the linkage between DRs. The link weights between nodes are defined by a matrix $\boldsymbol{W}$ that indicates DR similarity, where $w(r_i, r_j)$ denotes the similarity between r_i and r_j.

The traditional way to measure the similarity between textual content is based on single words as a vector space and cosine similarity measurement. It is insufficient to reflect the relevant concepts that are expressed using different words, such as the example of relevant issues “to improve print quality printhead” and “high drop generator density”. Therefore, in order to better model the similarity between the content of DRs, we intend to leverage neighborhood information in the DR network. We define a term relation $m(t_i, t_j)$ between term t_i and term t_j based on the local term relation $m_a(t_i, t_j)$ and the global term relation $m_b(t_i, t_j)$. The local term relation $m_a(t_i, t_j)$ is designed to model how the terms are related with each other within a space of a single DR. The global term relation $m_b(t_i, t_j)$ aims to model the relation between terms in the space of the DR network. t_i denotes a term in the term set T. $m_a(t_i, t_j)$ is defined under the assumption that if the terms often co-occur in the same layers within a DR, they are likely to be related, as shown in Figure 3(a) and Equation (1).

$$m_a\left(t_i,t_j\right)=\frac{1}{Z_a}\sum_{layer\in C}\left(\frac{1}{\left|position(t_i)-position(t_j)\right|}\right), t_i\neq t_j \tag{1}$$

It is measured by calculating the reciprocal of the positional distance between two terms, where *position*(t_i) represents the absolute location of term t_i of the document. *layer* includes the issue, solution and artifact layers of DRs in the whole document collection C. Z_a is the normalized factor, which is the summation of the reciprocals of absolute positional distances between all the terms.

$$m_b\left(t_i,t_j\right)=\frac{1}{Z_b}\sum_{t_i\in r_k,t_j\in r_l} e(r_k,r_l), e(r_k,r_l)>0, t_i\neq t_j \tag{2}$$

$m_b(t_i, t_j)$ is defined based on the assumption that if two DRs r_i and r_j have a linkage between them, then the terms appearing in r_i and terms appearing in r_j, which are in the corresponding layers, are probably correlated. To measure these relations, we use the neighborhood information of the DR network, as shown in Figure 3(b) and Equation (2). It is calculated by counting the number of linkages between DRs where the specified terms appear. We can also use the WordNet, which is a lexical database of English, to suggest the term relations based on their word senses. Z_b is the normalized factor. By integrating the neighborhood information, the term relation m is defined as shown in Equation (3):

$$m\left(t_i,t_j\right)=\begin{cases}\frac{1}{Z_m}\left[\alpha_a m_a\left(t_i,t_j\right)+\beta_b m_b\left(t_i,t_j\right)\right], & t_i\neq t_j\\ 1, & t_i=t_j\end{cases} \tag{3}$$

where α_a and β_b are the coefficients for parameter tuning, Z_m is the normalized factor. Based on the term relation, we define the similarity measurement between DRs, as shown in Equation (4):

$$w\left(r_i,r_j\right)=\frac{1}{Z_w}\left[\alpha_I w_{d_i,d_j}\left(I_i,I_j\right)+\beta_S w_{d_i,d_j}\left(S_i,S_j\right)+\gamma_A w_{d_i,d_j}\left(A_i,A_j\right)\right] \tag{4}$$

$w(r_i, r_j)$ the similarity between rationales r_i and r_j can be measured based on the ISAL structure by integrating the similarity of issue layer $w_{di,\, dj}(I_i, I_j)$ between rationales of document d_i and d_j, the similarity of solution layer $w_{di,\, dj}(S_i, S_j)$ and the similarity of artifact layer $w_{di,\, dj}(A_i, A_j)$. In order to support fuzzy search from different layers, we define the coefficients α_I, β_S and γ_A and allow designers to turn their values. All the coefficients in our case are defined as a real number in [0, 1].

$$w_{d_i,d_j}\left(x_i,x_j\right)=\frac{1}{Z_{wd}}[\alpha\sum_{\substack{t_k\in x_i,t_l\in x_j\\ t_k=t_l}} tw\left(t_k,d_i\right)tw\left(t_l,d_j\right)+\beta\sum_{\substack{t_k\in x_i,t_l\in x_j\\ t_k\neq t_l}} c\left(t_k,t_l\right)m\left(t_k,t_l\right)] \tag{5}$$

$$tw\left(t_k,d_i\right)=\frac{1}{Z_{tw}}tf\left(t_k,d_i\right)\times\left(1+\log_{10}\left(n/n_k\right)\right) \tag{6}$$

$w_{di,\, dj}(x_i, x_j)$ shows how to quantify the similarity between segment x_i belonging to document d_i and segment x_j belonging to document d_j, as defined in Equation (5). Segment x_i can be issue layer I_i, solution layer S_i or artifact layer A_i of document d_i. $c(t_k, t_l)$ is the number of co-occurrence that t_k occurs in x_i and t_l occurs in x_j. Z_w and Z_{wd} are the normalized factor. $tw(t_k, d_i)$ denotes the term weight for term t_k in document d_i, which is defined in Equation (6). $tf(t_k, d_i)$ is the term frequency of term t_k in document d_i, n is the number of sentences in d_i and n_t is the number of sentences that contain the term t_k. Z_{tw} is the normalized factor, which is the summation of term weights of all the terms in document d_i.

4.3 DR Retrieval Model Based on DR Network

As we discussed in Section 4.2, the common way of DR retrieval mainly focuses on finding the pieces of DR nodes containing a specific query by measuring the cosine similarity between nodes and the query. It neglects the relationships between DR elements that may well be helpful to suggest and rank the potential pieces of rationale according to designers' query. In our retrieval model, we intend to leverage the neighborhood information of the DR network for searching relevant rationales based on the fuzzy query from multiple aspects. Figure 4 shows the query input box of the rationale-based search. It allows designers to input query from issue, solution and artifact aspects, and it also permits designers to do fuzzy search by tuning the percentages of matching coefficients. If the coefficient is set 100%, it indicates the exact match of query words; otherwise, our retrieval model will conduct a fuzzy match by including terms that may be relevant to the query according to the term relation matrix.

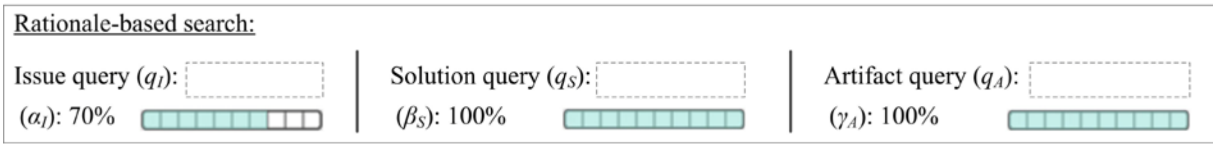

Figure 4. The Query Inputs Box for Rationale-based Search

Based on our DR network, we propose a graph-based algorithm to score each DR r_i according to the fuzzy queries, under the assumption that the nodes with high similarity are likely to have similar scores. We first process and expand the query through assigning initial score to each DR node r_i . We propose to use the concepts of statistical machine translation and the term relation to initiate the DR score. The initial $score(r_i)$ in Equation (7) is measured by the relevancy between r_i and the query q. $f(.)$ in Equation (8) denotes the similarity between fuzzy query and its corresponding layer of a DR, where x_i can be I_i, S_i or A_i, y_k is one of the coefficients α_I, β_S and γ_A.

$$score(q,r_i) = f(q,r_i) = f(q_I, I_i, \alpha_I) + f(q_S, S_i, \beta_S) + \gamma_A f(q_A, A_i, \gamma_A) \tag{7}$$

$$f(q_i, x_j, y_k) = \sum_{t_k \in q_i} p(t_k, x_j) = \sum_{t_k \in q_i} \sum_{\substack{t_u \in x_j \\ x_j \in d_j \\ 1 \geq m \geq y_k}} m(t_k, t_u) tw(t_u, d_j) \tag{8}$$

Next, the ranking process defined in Equation (9) starts to score each r_i by using the rationale similarity matrix $\boldsymbol{W}$. λ is the damping factor as in PageRank [20]. The nodes that are with more links will receive more scores from their neighbor nodes. The ranking process will continue until the summation difference of two successive iterations is lower than a given threshold.

$$score(q,r_i) = \lambda \frac{1}{|C|} + (1-\lambda) \sum_{r_j} w(r_i, r_j) score(q, r_j) \tag{9}$$

5 PRELIMINARY EXPERIMENTS AND RESULTS

Figure 5 shows the idea of prototype DR retrieval system. The first step is to collect the design documents as raw data from the possible sources. As for the external documents, such as patents and journal articles, we have implemented a crawling engine that allows designers to download their design documents of interest. In our study, we use patent documents as our research data, because the internal design documents, such as design reports and design records, are confidential and not publicly accessible. Patent documents as a valuable external resource for engineering design are quality data and accessible with critical rationale information.

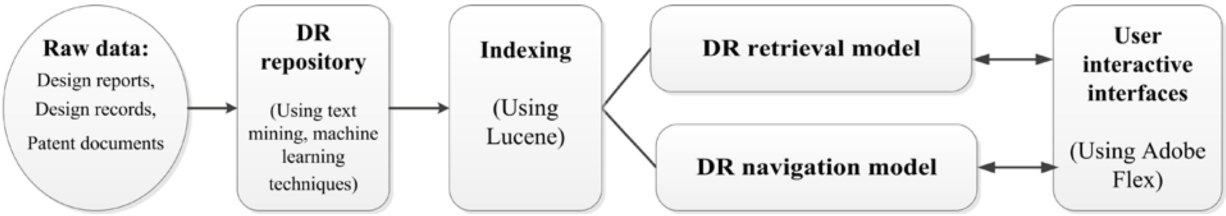

Figure 5. The Components of the Prototype DR Retrieval System

Each document in the raw data set is treated as a free text. Then the DR repository construction process is performed to transfer the design documents into a structural XML file as shown in Section 4.1. Particularly, the rationale information of a design document is extracted based on our ISAL model using text mining and machine learning techniques. Next, we generate the indexing term set *T* using our DP model. Based on the term set *T*, metadata and product classification, we use a tool Lucene [21] to index the XML. The DR retrieval and navigation model provide a retrieval strategy to exploring DR information. The user interactive interfaces provide functions that enable designers to search for and browse the DR network, which is implemented using Adobe Flex.

In the preliminary experimental study, we used twenty-four patents that are related to inkjet printer as sample data. We first evaluate the term set generated by our DP model compared with BoW for indexing purposes. In the DP modeling, we obtained all the single words and phrases that occur in at least one sentence as the term set. In BoW approach, all the single words in the sample data are selected. We calculate the averaged point-wise mutual information (PMI) of these two indexing term sets to measure the averaged strength of the semantic association among a set of indexing terms. The higher value of PMI indicates the stronger strength of semantic term association. In our study, we assume that if two terms appear in the same sentence, they tend to occur together.

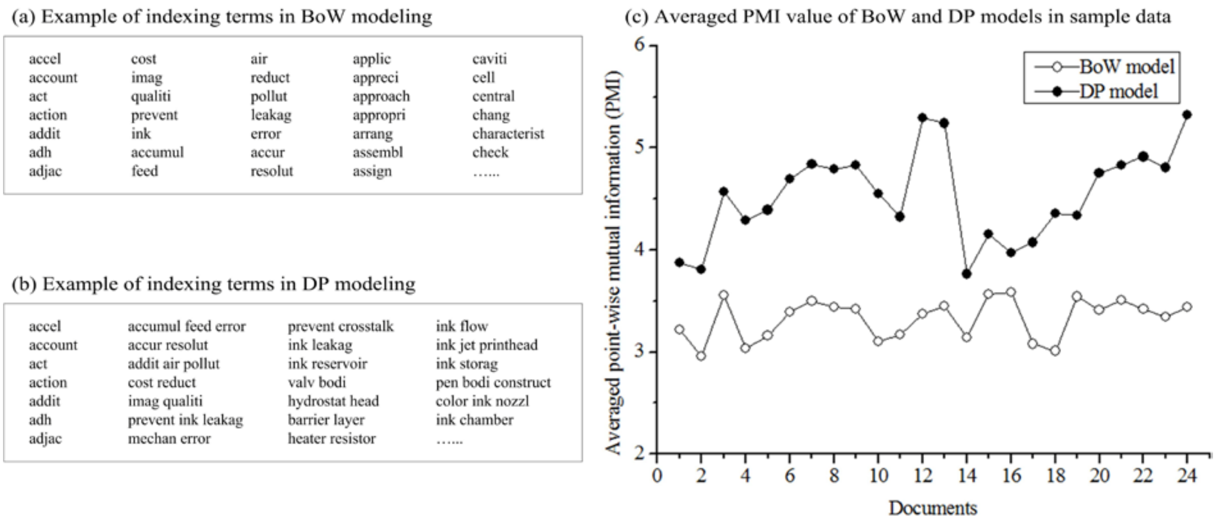

Figure 6. Examples of Indexing Terms by BoW and DP Modeling

Figures 6(a) and (b) show the examples of indexing terms generated by BoW approach and DP modeling respectively. We have noted that DP model can help to produce terms that designers may well use when describing an issue or an artifact component in a particular domain, such as “ink leakage” and “ink reservoir”. It is more readable compared with the single terms generated by BoW approach. From Figure 6(c) that shows the averaged PMI of BoW and DP model, it helps to reveal that the terms generated by our DP model is able to increase the sematic strength of the term association.

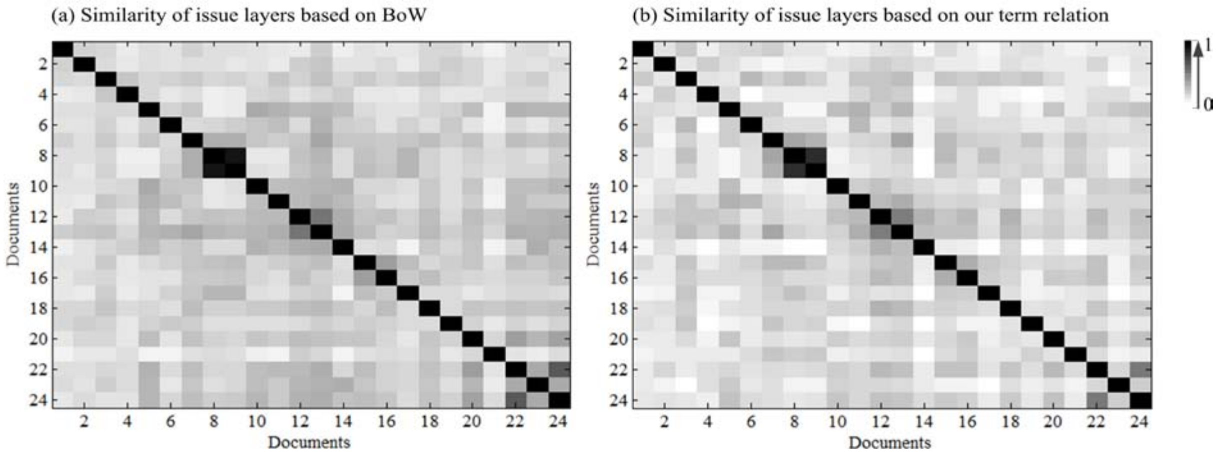

Figure 7. Similarity Measurement of Issue Layers by BoW and DP Modeling

Our next experiment illustrates the similarity of issue layers between the sample patents based on our DP model and term relations. For comparison, we also show the similarity based on BoW. Figure 7 illustrates the similarity results, where the darker color indicates the higher similarity value. The third columns of Figures 7 (a) and (b) indicate that our method of obtaining indexing terms helps to increase the similarity values between the third document and some other documents. The fourteenth columns shows that our method tends to decrease the similarity values between the fourteenth document and some other documents compared with that generated by BoW. These differences between similarities reveal that our approach is able to provide a possible separation between the individual layers in the documents to some extent. It can offer a better anticipation of relevant assessment between DRs.

In the third experiment, we use an example to show the use of neighborhood information for suggesting potential concepts. We create a scenario that a novel printer designer intends to search for the issue about “high speed printing” of “inkjet printer” as shown in the left hand side of Figure 8. The right hand side shows the potential terms that are related to the query words using the neighborhood information. For example, the DR content indicates that the issue of “high speed printing” is associated with some concepts such as “ink drop”, “nozzle” and “image quality”. The artifact components related to “inkjet printer” are like “inkjet printhead”, “ink cartridge” and “ink feed channel”. For a novel designer, the use of neighborhood information can provide suggestions on the correlations between concepts, so that they can gain relevant information about the design issues and the relevant design artifact components. Through this query keyword expansion, it also provides useful inputs for the retrieval process to search for the relevant rationales.

Issue query:	high speed printing	----→	Suggested terms:	ink drop, ink flow, nozzle, resolution, image quality
Artifact query:	inkjet printer	----→	Suggested terms:	inkjet printhead, ink cartridge, ink reservoir, ink feed channel, heater resistors, firing chamber.

Figure 8. An Example of Related Terms of a Given Query

6 CONCLUSIONS

The study of DR retrieval becomes an important issue in DR management as increasing amount of DRs is collected using computerized DR systems and stored in databases or digital libraries. In this paper, we have proposed a DR retrieval strategy by exploiting neighborhood information of the DR network formed and multi-dimension granular information. We first structure the rationale content in a document based on our previously proposed ISAL model. Then a DR network is constructed using neighborhood information. Based on DR indexing and its network established, a graph ranking based algorithm is brought forward to sort DRs according to their conceptual similarity with a given DR retrieval query. Some preliminary experiments are reported in order to validate the proposed strategy, including DR indexing, similarity measurement and concept based DR query expansion. More detailed benchmarking tests are being carried out and will be reported in our future work.

REFERENCES

[1] Regli, W.C., Hu, X., Atwood, M. and Sun, W., A Survey of Design Rationale Systems: Approaches, Representation, Capture and Retrieval. *Engineering with Computers*, 2000, 16(3), pp209-235.

[2] Burge, J.E. and Brown, D.C., Software Engineering Using RATionale. *Journal of Systems and Software*, 2008, 81(3), pp395-413.

[3] Bracewell, R., Wallace, K., Moss, M. and Knott, D., Capturing Design Rationale. *Computer-Aided Design*, 2009, 41(3), pp173-186.

[4] Tang, A., Jin, Y. and Han, J., A rationale-based architecture model for design traceability and reasoning. *Journal of Systems and Software*, 2007, 80(6), pp918-934.

[5] Wang, H., Johnson, A.L. and Bracewell, R.H., Supporting design rationale retrieval for design knowledge re-use. In *The International Conference on Engineering Design (ICED'09).* Stanford.

[6] Liu, Y., Liang, Y., Kwong, C.K. and Lee, W.B., A New Design Rationale Representation Model for Rationale Mining. *Journal of Computing and Information Science in Engineering*, 2010, 10(3).

[7] Kunz, W. and Rittel, H., *Issues as Elements of Information Systems*. (Center for Planning and Development Research, University of California at Berkeley, 1970).

[8] Conklin, J. and Begeman, M.L., gIBIS: a hypertext tool for exploratory policy discussion. *ACM Trans. Inf. Syst.*, 1988, 6(4), pp303-331.

[9] Shum, S., Selvin, A., Sierhuis, M., Conklin, J., Haley, C. and Nuseibeh, B., Hypermedia Support for Argumentation-Based Rationale. *Rationale Management in Software Engineering*, pp111-132, 2006).

[10] McCall, R.J., PHI: A Conceptual Foundation for Design Hypermedia. *Design Studies*, 1991, 12(1), pp30-41.

[11] de Medeiros, A.P. and Schwabe, D., Kuaba Approach: Integrating Formal Semantics and Design Rationale Representation to Support Design Reuse. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 2008, 22(Special Issue 04), pp399-419.

[12] Kim, S., Bracewell, R.H. and Wallace, K.M., A Framework for Design Ratioanle Retrieval. In *The 15th International Conference on Engineering Design (ICED'05).* Melbourne.

[13] Szykman, S., Sriram, R.D. and Regli, W.C., The role of knowledge in next-generation product development systems. *Journal of Computing and Information Science in Engineering*, 2001, 1(1), pp3-11.

[14] McMahon, C., Lowe, A., Culley, S., Corderoy, M., Crossland, R., Shah, T. and Stewart, D., Waypoint: An Integrated Search and Retrieval System for Engineering Documents. *Journal of Computing and Information Science in Engineering*, 2004, 4(4), pp329-338.

[15] Yang, M.C., Wood III, W.H. and Cutkosky, M.R., Design information retrieval: a thesauri-based approach for reuse of informal design information. *Engineering with Computers*, 2005, 21(2), pp177-192.

[16] Liu, S., McMahon, C.A., Darlington, M.J., Culley, S.J. and Wild, P.J., A computational framework for retrieval of document fragments based on decomposition schemes in engineering information management. *Advanced Engineering Informatics*, 2006, 20(4), pp401-413.

[17] Li, Z. and Ramani, K., Ontology-based design information extraction and retrieval. *Artificial Intelligence for Engineering Design Analysis and Manufacturing*, 2007, 21(02), pp137-154.

[18] Lim, J.S.C., Liu, Y. and Lee, W.B., Multi-facet product information search and retrieval using semantically annotated product family ontology. *Information Processing & Management*, 2009.

[19] Lim, S.C.J., Liu, Y. and Lee, W.B., A methodology for building a semantically annotated multi-faceted ontology for product family modelling. *Advanced Engineering Informatics*, In Press, Corrected Proof.

[20] Brin, S. and Page, L., The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 1998, 30(1-7), pp107-117.

[21] Hatcher, E. and Gospodnetić, O., *Lucene in Action*. (Manning Publications, 2004).

Contact: Ying Liu
National University of Singapore
Department of Mechanical Engineering

9 Engineering Drive 1
Singapore
Tel: +65-6516-7812
Fax: +65-6779-1459
Email: mpeliuy@nus.edu.sg
URL: http://serve.me.nus.edu.sg/liuying/

Ying is presently an Assistant Professor in the Department of Mechanical Engineering at the National University of Singapore. His research interests focus on Design Informatics – a multidisciplinary area that studies the representation, perception, processing, computation, communication, storage, retrieval and reuse of information and knowledge in the general context of design. He is a member of the Design Society, ASME and ACM.