DESIGN KNOWLEDGE DISCOVERY IN CASES

The Machine View vs. the Human View

CHIEH-JEN LIN
Department of Interior Design, Tainan Woman's College of Arts & Technology, Yung-Kang City, Tainan, Taiwan
jackyljz@seed.net.tw

MAO-LIN CHIU
Department of Architecture, National Cheng Kung University, Tainan, Taiwan
mc2p@mail.ncku.edu.tw

Abstract. In the previous study, we had applied the data mining techniques and ontology methodology to develop a keyword-based schema to extract and represent the implicit information within a case library, Case Base for Architecture (CBA). To improve the ability of our keyword-based schema on extracting and representing design knowledge within cases, we proceeded some experiments to understand the design's mental behaviors in extracting knowledge from cases. Through protocol analysis, we attempted to establish a knowledge discovery model of extracting design knowledge from cases, and to propose methods to apply this model to improve our keyword-based schema. Through collecting adjective keywords to re-structure our design dictionary, we attempt to make our system more sensitive to design knowledge, and more sensitive to user's intensions by extending the ontology of our keyword list.

1. Introduction

Design cases are often considered as the abstraction of previous design knowledge and are often becoming the impetus for design thinking and problem formulation based on analogy. Most of cased-based systems are established based on this view but mainly focuses on the collection and representation of cases rather than how to retrieve and reuse the design knowledge within cases. Current case libraries or case-based systems have less considered what kind of design knowledge is contained in cases and how to discover and represent those design knowledge. One major reason for this disadvantage is the technological difficulty to extract and index the implicit information within cases. Furthermore, the users may lack of appropriate association with the contents and representation of design knowledge within cases. Therefore, this paper is aimed to improve the reusability of case library based on keyword-based schema to extract design knowledge within cases, then we proceed a experiment to understand the design's mental behaviors in extracting knowledge from cases.

In the previous study, we had applied the data mining techniques and ontology methodology to develop a keyword-based schema (Figure 1) to extract and represent the implicit information within a case library, Case Base for Architecture (CBA). (Lin and Chiu, 2003) Meanwhile, we also visualized the semantic and ontological relations of keywords within cases to represent the explicit relations among cases. (Lin and Chiu, 2003) However, we found that the schema can represent the implicit information such as the similar design problems or solutions among cases, but it still difficult to convert implicit design information to explicit design knowledge in this way. The major problem is that users may be unfamiliar with the right keywords, especially for beginners or students who do not understand the meanings or significances of keywords. Another problem is that keywords extracted from contents of cases by the machine may be restricted and can't satisfy user's intensions or needs.

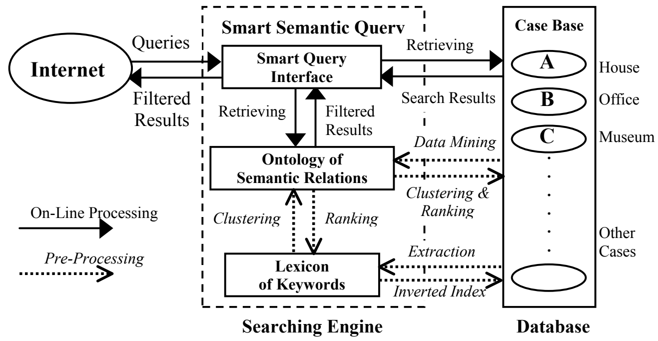

Figure 1. Framework of CBA

2. Experiments

To improve the ability of our keyword-based schema on extracting and representing design knowledge within cases, we proceeded some primary experiments to understand the design's mental behaviors in extracting knowledge from cases. In those experiments, students are requested to perform two tasks: (1) choose few keywords to represent the essence of selected cases based on keyword-based schema, and (2) use some keywords to find the most relevant cases as they can. Through the retrospective protocol analysis to validate the meaning and reasoning of keywords extracted by students, we try to establish a mental model of design knowledge extracting, students to help us to improve our keyword-based schema for design knowledge discovery and representation.

2.1. KEYWORD SELECTION EXPERIMENTS

First primary experiment is an attempt to understand designer's mental behaviors in choosing keywords to present their knowledge.

2.1.1. Keyword Selection without Relevant Cases

In the experiments, students first were requested to write down five most representative keywords to represent their impressions or knowledge for two issues: "hi-tech building" and "green building". Since the different background of our students that some students had more than 6 year experience in interior or architectural design since junior college, and some had only one and half years since thirst years of college, we found obvious different results varied with their different experiences and background knowledge.

For those students who had less experiences and knowledge, the results tended to be a random or literal interpretation of "hi-tech" or "green" based on their common sense, such as "computer", "electron", "internet", "green plants" and so on. For those experienced students, the results tended to be one of definite and specialized attributes or features about assigned issues, such as famous architects or cases name, building type or its style, design concepts or strategies and so on.

2.1.2. Keyword Selection from Relevant Cases

To reduce the difference in experiences and background knowledge, we sequentially introduced some cases of those issues, such as Pompidou Center by Rogers and Piano, Hongkong Shanghai Bank by Foster and so on. However, we did not attempt to describe those cases too detailed but just requested students themselves to collect and to review materials about those cases from books or Internet and furthermore asked students to choose keywords from those materials.

Because these experiments restricted students to choose keywords from those materials of relevant cases rather than to represent their own knowledge, students usually tended to choose most definite and characteristic features about the cases according their domain knowledge, such as case's function, type, style or construction system and so on. The results were very like the previous test by experienced students but more definite and specialized.

2.2. TEST SEARCHING OF CHOSEN KEYWORDS

After the keyword selection experiments, we sequentially attempt to test the searching and representing ability of those chosen keywords. Since there had no enough cases about assigned issues in CBA, we alternately chose three free, open and popular searching tools to proceed our experiments. Those tools are (1) the Google searching engine, (2) the "Great Buildings Online" database, and (3) the "archINFORM" database. We requested students to try three searching tools sequentially to test their own keywords selected from relevant case.

2.2.1. Test Searching on Google

There were about 4,930,000 results searched from Google search engine using keywords: "high", "tech" and "building". Obviously no one can review every page in this amount except the first several highest-ranking pages. However, as a searching engine for general purpose, Google still can provide some very important and useful reference pages in the higher-ranking results, such as high tech building catalogue in Great Buildings Online database. One of most interesting results when we tested Google is the very different results searched by similar keywords: "hi-tech" and "building". There are only 261,000 results in this searching and obviously different with previous results in the highest-ranking pages. Even so, it is very hard to explain there are any obvious semantic differences between "high tech" and "hi-tech" except differently spelling.

Most testing of using Google by our student were very like the primary "hi-tech" testing. Different keywords resulted in different results and more keywords resulted in less results. But there seem to have no other words that can find more precise cases than the keyword "high tech building" itself, and it also was very difficult to confirm which keyword can effectively reduce those irrelevant results but increase useful results because of too many results to review.

2.2.2. Test Searching in Great Buildings Online

Great Buildings Online is an open and abundant case library on Internet with more than 800 important architectural paradigms and we can easily to find out 23 cases in its "high tech" style catalogue through its hierarchy classifying system. However, we also found an interesting condition when we used this database. There were 23 cases found by the "hi tech" keyword, just as the results of "high tech" catalogue, but only 9 cases found by inputting the “high tech” keywords into the “style” field with the advance search function of this database. Furthermore, there was only one case with “high tech” found and 57 cases found by the “hi tech” keyword inputted into the “discussion” field in advance search, but most of those results were irrelevant to “hi-tech building” in our sense.

Except the “discussion” field, all the data fields contained in Great Buildings Online had been normalized in some degrees and its data had been summarized. Therefore the results by its advance search function usually cannot satisfy our student’s intensions. However, those keywords in some fields such as construction system provided our students useful clues in selecting their own keywords, such as “steel” and “glass”. But when our students tried to search again with those keywords, there were only several cases is similar with “hi-tech” style catalogue in 34 cases found by the keywords “steel” and “glass”.

2.2.3. Test Searching in archINFORM

Comparing with completely open database like Great Buildings Online, the archINFORM still need to register your email address for some detailed information, although it’s totally free, but it provides more profuse cases. There are more than 12000 cases, projects, architects and relevant publications in archINFORM, but all information in German and only partial in English. However, one advantage of archINFORM than Great Buildings Online is its will-organized hierarchy keywords lists that provide a useful tool for students and experts to browse relevant cases through its hierarchic catalogue. Although there were only 22 cases in the “high tech” keyword catalogue, but we could find 24 cases by its text search engine with the keyword “high tech” and we could not find any case with the keyword “hi tech” or “steel glass”.

The empty result by keyword “steel glass” came into our notice. We found that there are actually about 22 cases in “glass” keywords catalogue, about 10 cases in “steel” catalogue, and more relevant phrases such as “steel glas façase”, but the “glas” word lost one “s”. Therefore we thought that it might have some bugs in its search engine, and changed our searching keyword to “steel glas” with partial matching function. This time we can find about 70 cases from archINFORM with the new keyword “steel glas”. The results still were similar to our experiences in Great Buildings Online, because of too many irrelevant cases.

2.3. KEYWORDS SELECTED BY EXPERTS

It was found that our students tended to can only select out some relevant keywords to present partial essences of cases but were not enough to represent the concept of assigned issues because their limited experiences and knowledge. Therefore after the previous experiments, we tried to provided some keywords that we summarized from some textbooks, such as "exposed structure" or "exoskeleton", "articulated form", "industrial details" and "prefabricated parts" and so on for the "high tech building" issue to our students. However, we found those keywords still too abstract for our students to understand immediately without further explanations and some words cannot be found in those case libraries such as "articulated" and "exoskeleton".

However, although not all of keywords can be directly used to search relevant cases in those case libraries, but our students though that those keyword seems to more representative than those selected by themselves from the materials of relevant cases. Furthermore, after the introduction of keywords selected by experts, some experienced students can expend or reorganize their own keywords to help searching relevant cases in case libraries, such as attaching those adjective terms to their noun keywords.

3. Discussion

3.1. COMMON SENSE VS. DOMAIN KNOWLEDGE

To request students to select keywords to present specific issues of cases actually is to ask them to summarize and represent their knowledge about these particular issues. Therefore the results will be determined by student's experiences and knowledge about these issues.

In first state of keyword selections experiments before we introducing relevant cases, all students still are not very familiar with assigned issues, so they usually would attempted to interpret assigned issues with their own common sense by their semantic relation, and thus usually firstly focused on the literal meaning of partial issues that they can understand immediately such as "hi-tech" and "green" and ignored the "building" issues. However, for some experienced students, they will further try to connect their common sense with their domain knowledge to find more meaningful representations.

In the further state after prompting relevant cases, because the materials of relevant cases that students reviewed became more definite in special knowledge domain, so students tended to detect which parts in materials were most fitting with their domain knowledge, such as which attributes will result in those cases to be more "hi-tech" or "green" than others. Therefore our students found some common attributes in those cases such as similar construction system, exterior materials and shape, and thus detected them as the necessary properties of assigned issues.

However, because constrains of the retrospective protocol analysis, we presently cannot validate the details processes of our students how to detect the semantic relations among assigned issues, their common sense and domain knowledge yet. Bur the semantic relations seems to be one of the most important clues in these processes, but all of detailed processes still needs more investigation and analysis.

3.2. NOUN KEYWORDS VS. ADJECTIVE KEYWORDS

In first stage of keyword selection experiments, students were not restricted to use any special morphological terms as their keywords. Therefore there were various terms in student's selections rather than only noun. However, perhaps inspired or affected by experiences, students began to more notice noun keywords than other after testing Great Buildings Online database. But after testing the advance searching of Great Buildings Online with these noun keywords, for examples the "steel" and "glass", students began to find those keywords seem to be too general and usually resulted in many irrelevant results. Based on the experiences of using Google, thus some experienced students began to try to attach more detailed words, such as adjective terms like "exposed" and "structural", to try to reduce irrelevant cases. Nevertheless, these attempts usually cannot be satisfied because the normalized contents in Great Buildings Online and archINFORM database.

Most of information systems usually consider that noun terms are more significant in representing abstract concept or realistic things than other and usually tend to select noun terms as their keywords for searching purpose. (Baeza-Yates, and Ribeiro-Neto, 1999) However based our observation in those experiments, the phrases with adjective terms should be more significant for our students than single noun terms in representing indistinct issues such as "hi-tech" or "green building" issues.

3.3. DATA MAPPING VS. INTENSION PRESENTATION

In those experiments testing Google search engine, some experienced students took an upside-down strategy to overcome the amount of retrieved results. Based on their past experiences using Goolge, they firstly collected a set of results directly using words of assigned issues such as "high tech" or "green building", and then added more words to search again in this set to reduce the irrelevant results. Those additional words were chosen by their common sense or domain knowledge as we addressed before.

However, we found that some students who had less experience seemed to image their searching process like to find more definite words in order to interpret those abstract issues to an innocent beginner, i.e. the search engine machine, and expected this machine to understand their intensions. But other more experienced were more careful in repeatedly selecting and testing their keywords because they knew that searching machine just checked those inputted keywords whether in the contents of web pages or not.

The “high tech” and “hi tech” testing in Google evidently illustrated this difference between machine and human. Goolge obviously saw “hi” and “high” as two different words with different meanings, but we usually expected that machine should be smart enough to avoid this difference. We thought this difference was not only caused by technological constrains, but also by the different concept and strategy between human and machine.

The strategy of searching engines is based on data mapping of information retrieval techniques. Through comparing the queries from users with the contents in data set and ranking retrieved relevant results, searching engines usually only cares about exact words rather than the intensions of users. However, queries from users usually do not an irrelevant collection of words but a representation of user’s searching intension. To understand user’s intension should be more beneficial than just matching words although this idea is very difficult in technology view.

4. Conclusion

Although our experiments were still very primary and there were many issues need more time to investigate. It is found that design knowledge extracted from cases may vary with use’s experience, knowledge and situations they occurred. Same user may learn or discover different knowledge from same cases at different time or situations. Users’ experience and knowledge seem to be a critical factor in extracting knowledge from cases. Key questions are raised such as: How does a case-based system help user to extracting knowledge? How does a case-based system extract knowledge from cases more efficiently and effectively?

In conclusion, this paper has observed some phenomena in designers’ mental behaviors of extracting knowledge from cases. Then we attempt to establish a knowledge discovery model of extracting design knowledge from cases to help us to improve our keyword-based schema in CBA. (Figure 2)

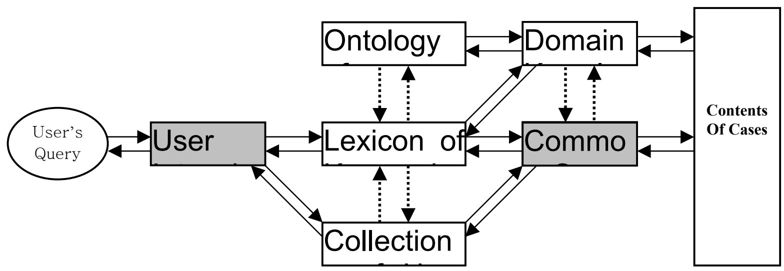

Figure 2. Knowledge Extracting Model in CBA

4.1. COMMON SENSE AGENT BASED ON DOMAIN KNOWLEDGE

In our issue-oriented experiments, we found that it was very difficult for our students to find an appropriate and definite definition of our assigned issues without any help. Actually we though the concept of “high tech” or “green building” seems to be a indistinct and relative issues about the constructional, energy saving and environmental protection technology of building. Although the standard of “higher” or “greener” technology may be very clear for experts but still very indistinct for beginners. Therefore our students would try to estimate which attributes or parts in relevant cases were “higher” or “greener” than or similar with their experiences or domain knowledge.

For example, experts seems to consider the steel and glass structure as more “hi tech” and “greener” than concrete system because of its more industrial and recyclable properties. But our students obtained similar results from relevant cases by inductive inference rather than by understanding their properties in domain knowledge. Therefore we think that there should be a common sense agent based on domain knowledge to detect the properties and attributes of domain knowledge in cases, such as automatically collecting the relevant keywords about the attributes of steel structure systems, to help users to extract useful knowledge from cases and to build up their common sense of that domain knowledge.

4.2. EXTENDING KEYWORDS TO ADJECTIVE TERMS

For detecting the properties and attributes of domain knowledge, we though that we should pay more notice to adjective terms for presenting relevant attributes rather than only focus on noun terms to present abstract concepts or definite knowledge. For example, those words such as “industrial” and “prefabricated” for “hi tech” issue, and “recyclable” and “sustained” for “green building” issue, should be more explanatory and representative for the properties and attributes of “steel structure system” about different issue. Therefore we will try to extend and re-structure out design dictionary by more adjective terms collected from user and from contents of cases to make our system more sensitive to the attributes of design knowledge contained in cases.

4.3. INTENSION AGENT BASED ON COMMON SENSE

Once we can establish the common sense agent and make our system more adjective to the attributes of design domain knowledge, then we should be able further to extend the ontology of our keyword list by the ontology methodology. Thus we will also attempt to build up an intension agent to make our system more sensitive to user's query intensions to help user to find more relevant cases by their intension rather than keyword mapping.

However, all of those ideas in our knowledge-extracting model still need more investigations and we will attempt to implement and integrate them in to our CBA system in future works.

References

Baeza-Yates, R. and B. Ribeiro-Neto.: 1999, *Modern Information Retrieval.* New York: Addison Wesley.Maher, M. L., Balachandran, M.B., Zhang, D.M.: 1995, *Case-based Reasoning in Design*, Lawrence Erlbaum Associates, Inc., New Jersey.

Chiu, M.L., C.J. Lin, T.S. Jeng and C.H. Lee.: 2002, Re-Searching the Research Problems in CAAD: Data Mining in i-CAADRIA. *Proceedings of CAADRIA 2002,* Kuala Lumpur, Malaysia: Prentice Hall, pp.31-38.

Lin, C.J., Chiu, M.L., 2003, Information Mining within a Case Library: Visual Links of Correlation among Cases, Proceedings of 8th International Conference on Computer Aided Architectural Design Research in Asia (CAADRIA), Bangkok, Thailand, Rangsit University, pp.249-261.

Lin, C.J., Chiu, M.L., 2003, Smart Semantic Query of Design Information in A Case Library, Digital Design: Research and Practice, Proceedings of the 10th International Conference on Computer Aided Architectural Design Futures (CAAD Futures), Tainan, Taiwan, Kluwer Academic Publishers, pp.125-135.

Martens, B., Turk, Z., 2002, Digital CAADRIA-Proceedings: Retrospective Analysis of Content, *Proceedings of CAADRIA 2002,* Kuala Lumpur, Malaysia: Prentice Hall, pp.23-30.

Riesbeck, C. K.: 1996, What Next? The Future of Case-Based Reasoning in Post-Modern AI. In *Case-Based Reasoning: Experiences, Lessons, & Future Directions,* ed. David B. Leake, Cambridge: the MIT Press, 1996, PP371-388.

Simoff, S.J. and M. Maher, 1998, Ontology-based Multimedia Data Mining for Design Information Retrieval. *Preceedings of ACSE Computing Congress.* Cambridge.

Witten, I. H. and E. Frank.: 2000, *Data Mining – Practical Machine Learning Tools and Techniques with JAVA Implementations.* San Francisco: Morgan Kaufmann.