Information and IN-formation
Information mining for supporting collaborative design

Mao-Lin Chiu$^{a,*,}$, Ju-Hung Lan$^b$

$^a$Department of Architecture, National Cheng Kung University, No. 1 University Road, 701 Tainan, Taiwan
$^b$Department of Architecture, Kao Yuan Institute of Technology, Kaohsiung, Taiwan

Abstract

Collaborative design has become a research paradigm in design studies. To make effective collaborative design, an information service mechanism for helping collaborators to access related information of specific design situation is getting important. This paper presents an approach of applying data mining techniques to reveal information patterns for managing collaborative design information. A visual interface of linking design information based on revealed patterns are presented and issues are discussed.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Information; Data mining; Collaborative design; Web-based system

1. Introduction

Collaborative design has become a research paradigm in design studies ([1–4]). While the establishment and use of computer-supported collaborative work (CSCW) systems can ease the design process, users have to manage vast amounts of information available in the systems. An effective searching mechanism to help users find desired information is becoming important.

Data mining in database is to find previously unknown and potentially useful knowledge from vast amount of data. The main task in the processes is to discover meaningful patterns among data [5]. Therefore, data-mining techniques is adopted to discover meaningful information patterns from collaborative design projects. This paper presents an approach to develop a web-based information management system to support collaborative design by applying data-mining techniques. The concept, system implementation, and discussion are presented.

2. Design information

How design data are generated or collected “into” the design “formation” process is the key factor for information processing. The typical design information in a collaborative design process includes design
documents, drawings, communication messages, charts, 3D models, etc. [6]. In this research, design data including text or non-text information are considered as raw data in the process. In general, most design documents are not standardized and without keyword features. The variety of information stored in the backend database of a collaborative design system could be classified as structured and ill-structured information. The structured information, for instance meeting briefs or memos, usually provides tree-structure classification functionality of design information in accordance with specific design information group. The ill-structured information, for instance message board data, usually are sequential records in accordance with timing of information posted. It is necessary to abstract relations among both structured and ill-structured information for further applications.

When data are organized into design information for communication purposes, a specific design situation related to participants, timing, places, tasks, or rationales occurred. Therefore, a 6W (who, when, where, what, how, and why) approach is applied for associative linking related information. For the practical reasons, the collected information is classified into three levels: project, group, and individual. The information created in the process can automatically be organized into sub-data sets by the 6W-based classification, i.e. who are involved, or when, why, what, where, and how information are occurred. Each participant then can access the information from the individual, group, or project point of views. For facilitating design services, the correlation of 6W-based information can be developed as content-awareness mechanism as shown in Fig. 1.

3. Information mining

How to discover the hidden information service patterns regarding the 6W-related information is critical for this research, particularly to discover the classification and association patterns of collaborative design information. Our intention is to manage collaborative design information with a semantic structure by the classification and association patterns discovered. Then we could provide a situated information service for collaborators during design processes. The data-mining techniques are used to explore the hidden patterns among 6W-related information. The framework of information mining of this research is shown in Fig. 2.

Information classification is generally a top-down process, while information association is a bottom-up process. In terms of classification, tree-structure with attributes setting can be used to represent the categories of information belonging to, which is design situation oriented and serves as the information classification patterns. The attributes setting of each information node in the network-structure is regarding the 6W features mentioned above, which serves as the information association patterns in this research. The overall relationship of information patterns in this research is shown in Fig. 3.

The important steps of information mining include data pre-processing and discovery of classification patterns [7]. Data pre-processing is to clean data collected from design projects by filtering out useless data and encoding useful data into an integrated database. The task of filtering data is to avoid unsuitable design information collected from design projects. On the other hand, the task of encoding data is a mapping process of transferring
original design information into a simplified but meaningful format for indexing purpose. This is especially useful for dealing with the ill-structured design information such as drawings, charts, or communication messages collected from design projects. In this research, the attributes settings regarding the 6W are also encoded and assisted by domain knowledge. For instance, a pre-defined design terminology regarding the information classification and association, shown in Table 1, is used to encode the information related to opening design situation based on domain knowledge. This is a direct approach to get a reliable encoding result. Furthermore, it is an easier way to deal with the issues of synonyms, double-entendre, or different language at the initial stage.

After establishing the raw data set, this research attempts to develop classification patterns by building the hidden classification terminologies of design information. The statistic techniques and domain knowledge are used to produce the summary analysis of collaborative design information to find frequently used design terminologies with information classification capability. In addition, the collaborative design processes, which is design situation oriented in general, actually reflect the hierarchical and semantic relationship among design information. Therefore, we could classify design information regarding the different design situations with a tree-like semantic structure by mining the classification terminologies.

To develop an information association mechanism with a network-like semantic structure to associate
design information, this research first enriches the original design information with attributes settings regarding the 6W features. The 6W actually reflect the network-like semantic structure among design information. The design information could be associated with each other by tracing the 6W attributes. Meanwhile, this research also uses data-mining techniques to find the frequent terminologies of associative keywords to enhance the network-like semantic structure of the association mechanism.

The results of information mining in the above studies indicate that: (1) the classification mechanism could be used to develop an information service system that is able to classify design information regarding the specific design situation by a tree-like semantic structure; and (2) the association mechanism, on the other hand, could be used to develop an information service system that is able to associate design information regarding the 6W features by a network-like semantic structure.

4. System prototype

To support collaborative design activities, a groupware prototype of collaborative design system, named as Distributed Environment for Collaborative Architectural Design (DECADE) is implemented (Fig. 4). The research uses DECADE to collect collaborative design information as the platform for mining the information service patterns. To develop a CSCW system with classification and association capabilities, two main tasks are: (1) to develop an information classification mechanism with a tree-like semantic

![Fig. 4. The DECADE system prototype.](image-url)
structure; and (2) to develop an information association mechanism with a network-like semantic structure.

Data pre-processing of this research is completed by tablet analysis with the collected design information followed by database techniques to construct the raw data set for information mining. Text-based information can be converted into XML files. Non-text information such as images is indexed by the labels of information.

5. Visual links of design information

This research particularly focuses on the visual interface of linking text information. Traditional information management systems use keywords to search available information in data set. The result of the search is often listed with detailed and separated information. However, this kind of searching mechanism is difficult to deal with vast amount of design information [10]. It cannot find such as synonyms, double-entendre, or a word with various meanings. Nor it can provide the searching ability to find associated design information regarding the specific design situation. The capability of this kind of searching mechanism is static, passive, and lack of design knowledge supports. On the contrary, associative links enable users to connect information into a network of logical association with 6W approach ([8,9]). Furthermore, visualization of association becomes more feasible and important on the web searches now.

It is important to organize design information among inter-disciplinary knowledge domains for developing the effective information profiles. The analysis is based on the uses of the prototype system. When the information is gradually increased during the design process, users are more critically dependent on the links for associative reasoning. Consequently, the growing information will be re-organized and subdivided into time, theme, or issue-related sub-data sets. The research findings indicate that the visual links can assist users to find desired information with computer assistance. The retrieved information will support design decision by direct or indirect links. The information classification and association in a design situation is shown in Fig. 5. The visual links indicate the hidden associative information of interior design related with this design situation.

6. Empirical studies

To demonstrate the effectiveness of the system, the research conducts the empirical studies of two collaborative design projects: (1) the library project and (2) the free-form facade project. In both projects, the participants are distributed in different locations, and rely on the design information for communication and design development. For comparison purposes, the DECADE system is applied to the library project, while the free-form facade project does not use the
system and relies on the traditional way to communicate the design information.

The information created in the library project is linked to the system, and DECADE can automatically organize the information into sub-data sets by the 6W-based classification. Therefore, the participants can get content-awareness information services in DECADE to refer to a specific design situation during the design process. On the other hand, the information collected in the free-form façade project is chronologically putting on files regarding the specific design stages by a team-oriented but manually working flow. Meanwhile, there is no automatic way to structure the information and provide effective information service from the individual, group or project point of views. Without the system assistance, the participants usually spend a lot of time to find the related information they needed to a specific design situation.

The research starts information mining from few similar design situations such as opening design, façade design, roof design to associate envelope design information. Fig. 6 illustrates partial visualization results of information classification and association in the library project. The information in the opening design branch links to the interior design branch by the shading terminology, lighting terminology, and 6W features. The visualization results provide great potentials about discovering hidden information service patterns. Briefly, these two mechanisms provide the capabilities to develop an integrated and active information service system that this research intends to achieve.

7. Discussion

The above analysis provides the foundation for the following discussion, including design ontology dictionary, pattern approach, the volume of information, visual linkage of information, and information provision for supporting CSCW.

7.1. Design ontology dictionary

The establishment of the design ontology dictionary is subject to domain knowledge, expertise, and
applications, while the dictionary is also constrained by how information are generated or prepared. In this research, we built the ontology dictionary by mining collected design information from typical design situations such as opening design or facade design. However, in comparing with the professional dictionary that has about 10,500 terms, our dictionary containing about 2000 terms can be restricted because the information collected is based on few situations abstracted from two projects. The ideal dictionary could be more comprehensive but selective. Therefore, it is necessary to collect more situations for testing the usability of dictionary.

7.2. Discovery of patterns

The information collected in this research is design situations related with 6W features, which are the keys to discover the classification and association patterns. In the design process, the information collected can be classified into 6W-related patterns, i.e. who is asking question, what information is related, when information is created, how information is generated and where the information is located. The research found that users are first concerned about who and when information can be provided in collaborative design (Fig. 7). An effective information service mechanism based on 6W features to detect situations should be provided for supporting collaborative design. Our study found that 6W features can be complex but at least 3W (who, when, what) to assist situation detections will be feasible. Therefore, the information provision for supporting CSCW could be more effective during collaborative design processes.

7.3. The volume of information

In order to access the most related information, information can be classified and associated by building types, design phases, building systems, sub-domain or expertise. In the free-form façade project, the information generated and collected are less because the design task is specific and functional. On the contrary, in the library project, a large amount of information is generated by the participants and the volume of information will increase rapidly during design processes. How to classify and associate a vast

![Diagram of design domain]

Fig. 7. The information of building envelope design required for collaborative design decision.
amount of information in a CSCW system becomes an important task. It is found that users are normally interested in some key information. In order to maintain the ideal volume of information appeared, information can be classified and associated by a limited amount of nodes or linkages.

7.4. Visual linkage of information

The visual linkage of information can provide a visualization effect to associate design information. The information nodes and links provide the hints of information classification and association. In this research, the information node is encoded with design terminology representing specific information classification. Therefore, users could get more effective information service by this kind of visualization technique. However, it requires more sophisticated interface when the number of nodes or information is growing. For further studies, the links among information regarding the 6W-based associative capabilities that can be highlighted by various colors, thickness, or representations.

7.5. Information provision for supporting CSCW

Although the finding of system usefulness is still preliminary in the study, the information mining is promising for supporting collaborative design in the early design stage. While it is difficult to distinguish the difference of information needed at the project, group, or individual level, frequently used information are important to the persons who create the information as well as the correspondent. It is found that users often have different levels of needs on each aspect (W) or issue. Therefore, it is desirable to define different levels of classification and association by users for various purposes, for instance, to choose partial important issues by using first 3W (Who, When, What) for design progress review, and then expanding to another 3W (Why, hoW, Where) for problem definition or decision-making. In a complex project, it is more feasible to use one W (Who) first, followed by design situation consideration, to classify and associate design information. In this manner, the system can facilitate users to handle vast amount of information on hand.

8. Conclusion

In current CSCW systems, an information service mechanism to access related information of specific design situation is getting more and more important. This paper presents an approach of applying data-mining techniques to reveal the information classification and association patterns for managing collaborative design information. With the information service mechanism developed by this research, the collaborators could get more situated information service regarding a specific design issue, person, event and stage.

To expand the capabilities of current CSCW systems, the study proposes a visual interface for associative reasoning by information mining techniques. The visual interface of linking design information shows great potentials to discover associative and meaningful information by the 6W features. The collected patterns from empirical studies can be developed into a content-sensitive query mechanism. In the future development, adding the learning ability could enhance the performance of information mining within the system.

Acknowledgements

This research is supported by National Science Council in Taiwan, NSC91-2211-E-244-005.

References


