Web-based Knowledge Database Construction Method for Supporting Design

Kiyotaka Takahashi 11, Aki Sugiyama 11, Yoshiki Shimomura 11,
Takeshi Tateyama 11, Ryosuke Chiba 11, Masaharu Yoshioka 2), Hideaki Takeda 3)

1) Faculty of System Design,
Graduate School of System Design,
Tokyo Metropolitan University
6-6 Asahigaoka, Hino-shi, Tokyo,
191-0065, Japan
+81-42-585-8425
takahashi-kiyotaka@sd.tmu.ac.jp

2) Graduate School of Information Science and Technology,
Hokkaido University
Kita 14 Nishi 9, Kita-ku, Sapporo,
Hokkaido, 060-0814, Japan
+y81-11-706-7107
yoshioka@ist.hokudai.ac.jp

3) National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku,
Tokyo 101-8430, Japan
+y81-3-4212-2543
takeda@nii.ac.jp

ABSTRACT

Recently, designers have been required solving comprehensive problem becoming greater and more complicated. In relation to this background, we have proposed Universal Abduction Studio (UAS); a computer environment that synthetically supports creative design. However, it is difficult for a designer to manually acquire multiple domain knowledge required for UAS. Therefore, we propose a Web-based knowledge database construction method for supporting design by UAS in this paper.

Categories and Subject Descriptors
H.2.8 [Information Systems]: Database Applications – data mining.

General Terms
Management

Keywords
Creative Design, Knowledge Extraction, Analog, Ontology

1. INTRODUCTION

In recent years, the number of problems related to artefact design has increased for many reasons: the environmental problems, the magnification and complication of artefacts and the diversification of consumers’ values. Because of these issues, designers’ roles and responsibilities have increased. In order to overcome these problems, the authors have developed a computer environment called Universal Abduction Studio (UAS) [10] to support creative design based on new knowledge generation processes in creative design; UAS realizes step-by-step knowledge extension by integrating different domains’ knowledge groups. In the knowledge integration, referring to a large amount of knowledge is effective. Therefore, designers need to store a huge amount of knowledge groups in multiple domains in the UAS design knowledge database. However, it is hard for them to acquire knowledge groups from multiple different domains because designers need to manually extend the knowledge stored in UAS. Moreover, we previously proposed a method to convert design knowledge data described in natural language into data with a format for UAS [12]. Though its processing accuracy was not so high, designers need to manually correct design knowledge data before conversion.

In this paper, for automatic and efficient knowledge construction, we propose a Web-based database construction method. For this purpose, we first propose a method of acquiring knowledge on the Web; this automatically identifies design knowledge, and extracts sentences with useful information for design knowledge (hereafter, we call this “design knowledge information”). Second, we propose a method to reduce information that is useless for designers from the design knowledge information database. Third, we propose a method to convert the acquired design knowledge information into available knowledge expression data for UAS. We introduce the concept of UAS in section 2. In section 3, we propose a Web-based design knowledge construction method. In section 4, we verify the effectiveness of the proposed method. In the final section, we conclude the paper.

2. UNIVERSAL ABDUCTION STUDIO

![Figure 1. System architecture of UAS [10]](image-url)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. iiWAS2008, November 24-26, 2008, Linz, Austria. ©2008 ACM 978-1-60558-349-5/08/0011 $5.00
Universal Abduction Studio (UAS) is an abduction-based creative design support system. Figure 1 shows the system architecture of UAS. A workspace, a knowledge base group and a knowledge integration module group compose UAS. The fundamental concept of UAS requires a unified knowledge description among various domain knowledge bases. In UAS, rule-type knowledge expressed in an If-Then form is used as a knowledge representation form (hereafter, we call this “If-Then type design knowledge”). Moreover, it is necessary to add information about valence grammar and case grammar for UAS in the design knowledge data. Therefore, we proposed an automatic document tagging method [12] in order to convert the design knowledge described by natural language into graph geometry as UAS knowledge. This method delimits the design knowledge to If-Parts and Then-Parts at the part including 10 patterns of words and Parts-of-Speech extracted from existing design documents.

3. WEB-BASED DESIGN KNOWLEDGE DATABASE CONSTRUCTION METHOD

3.1 Procedure of the Proposed Method

The authors use Web information as the source for design knowledge in order to automatically and efficiently acquire multiple domain knowledge groups. Especially, in this paper, we use press releases on the Web as the source for design knowledge because press releases are highly reliable and include design information such as demands of customers who are targets for new products, the function to satisfy the demand, and the mechanism for achieving the function. Figure 2 shows Web-based design knowledge database construction method for UAS.

1. Acquire Web information with Websteemer.
2. Extract the design knowledge information from the Web Information DB.
3. Reduce the useless information for design in the design knowledge information DB.
4. Convert the extracted information into knowledge representation for UAS.

Figure 2. Web-based design knowledge database construction method for UAS

The details of the procedure are described as follows.

1. Acquire Web information including design knowledge using Webstemmer [5], a Web crawler for news sites, and store it in the Web information database.
2. Acquire the design knowledge information from the Web information database from Web information DB using a method of extraction of design knowledge information from the Web information DB.
3. Reduce the useless information for design in the design knowledge information database using a filtering method.
4. Convert the design knowledge information into design knowledge for UAS using automatic conversion methods for knowledge representation, and store it in the design knowledge database for UAS.

3.2 A method of extraction of design knowledge information from the Web information DB

In this section, the authors introduce the term design tag defined in this research. Moreover, the authors explain how to create a design tag, and how to automatically apply the design tag to Web information. The authors define a design tag is the tag that is tagged to the word related to design knowledge. We use it as a label for the efficient acquisition of useful design information from Web information. The design tag operates as the measure which shows how many concepts composing design knowledge the acquired design information contains, and we assume that a sentence including some design tags is useful design information. Therefore, we extract paragraphs that include design knowledge information sentences added design tags. In this research, the concept of the design tag is based on “The broader concept of the feature word of an article”, such as a physical phenomenon, a property in GoTaikei [1]. The authors define a feature word is “A word existing with high frequency”. In this research, a feature word is chosen by applying the TF (Term Frequency) method to nouns, verbs, adjectives and adjectival verbs. Next, we explain the method for creating the design tag set. First, we acquire multiple pieces of information from the Web and extract candidates for the design tag set from Internet. In this paper, we use press releases published in Nikkei Net [6] as an example of press releases about products. For example, we acquire articles on the four different domains of beverages, digital cameras, electric power equipments and Web services from Nikkei Net, one by one, to avoid bias in the knowledge domains. The design tag set is created using the following procedure. Figure 3 shows an example of creating a design tag.

1. Execute a morphological analysis of sentences in the press releases and a frequency analysis for each morpheme.
2. Extract the concepts of the extracted feature words in procedure 1 based on GoTaikei.
3. Abstract the extracted broader concept in procedure 2 to make it “a concept which does not depend on knowledge domains”, and define the highest concepts as design tags.

Figure 3. Concrete example of the method for creating a design tag set

Procedure 3 makes it possible to acquire useful information for the design regardless of the domain of the press release. In addition, we applied procedure 3 to the concepts of all the listed feature words in procedure 1 and created multiple design tags. For the frequency analysis for feature word extraction, we use a morpheme frequency analysis program developed by the authors. We adopted the name of a concept in GoTaikei as the name for a design tag and we added an initial D, meaning “Design”, to their heads. If a term had multiple broader concepts, we allocated the appropriate broader concept to the term by considering the contextual meaning of the term subjectively. Table 1 shows the design tag set created in this research. Next, we explain the automatic tagging method for design tags. First, designers execute a morphological analysis by using Chasen [11] for each article acquired by Webstemmer. Second, they match each morpheme in the article and each term that should have a design tag added among the terms in GoTaikei and add
design tags to the morphemes matched to them. In this thesis, objects that should be tagged are all terms that belong to the subordinate concept of each tag. Because prototypes and inflections exist in Japanese verbs, we perform the matching by using prototype information for vocabularies outputted as morphological analysis results by Chasen. Finally, all the paragraphs, including the sentence tagged by the above-mentioned procedure, are extracted. Each task in the automatic tagging method for design tags is automatically completed by a computer (processing with the grep command of Linux or an existing substitute program).

Table 1. Design tag set created in this research

<table>
<thead>
<tr>
<th>Japanese sentence</th>
<th>Design tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>-</td>
</tr>
<tr>
<td>Animal</td>
<td>-</td>
</tr>
<tr>
<td>Physical phenomena</td>
<td>-</td>
</tr>
<tr>
<td>Property</td>
<td>-</td>
</tr>
<tr>
<td>Right</td>
<td>-</td>
</tr>
<tr>
<td>Approved</td>
<td>-</td>
</tr>
<tr>
<td>Remove</td>
<td>-</td>
</tr>
</tbody>
</table>

3.3 Filtering Method

In the information included in press releases related to new products' release information, most of the sales information and corporate information is less likely to give designers new insights for design. Therefore, if we process the press releases that include this information using the proposed method, a large amount of useless information will be included in the design knowledge database. As a result, computational complexity and processing time for the inference of UAS will increase pointlessly. Consequently, we propose the following filtering method to reduce information that is useless for design in the design knowledge database. We assume that the following six named entity tags defined by IREX [9], <ORGANIZATION>, <PERSON>, <LOCATION>, <DATE>, <TIME>, and <MONEY> are added to sales information that seems not to be useful for design. These tags are called "Reduce tags." Subsequently, we explain the procedure of the filtering method. First, designers create articles that include the named entity tags using Cabocha [3] with the named entity analysis program. Next, they reduce the paragraphs including reduce tags.

3.4 Automatic Conversion Method for Knowledge Representation

The following two kinds of problems exist in the automatic document tagging method.

- Erroneous handling of one sentence that include two or more design knowledge, but design knowledge information often includes two or more design knowledge.
- Low processing accuracy. This system cannot correctly process a lot of If-Then type design knowledge data.

To solve the above problems, we propose two methods as automatic conversion methods for knowledge representation.

3.4.1 The Method to Divide Design Knowledge into One Design Knowledge in One Line

The automatic tagging method for design tags acquires including design knowledge information. However, we think it is often that the causal knowledge is described in one sentence. Therefore, we think that it is appropriate to divide design knowledge information into one design knowledge in one line. Hence, we propose a system that can place a period "." to punctuate the sentence and automatically divide design knowledge information into multiple sentences with the regular expression. This system is implemented in Ruby.

3.4.2 The Improved Method of the Automatic Document Tagging System

We describe related work on If-Then delimitation as follows. Inui et al. analyzed newspapers, and categorized words that showed a causal relation as causal markers. In addition, they proposed a method for acquiring cause and effect knowledge including "tame (one of the conjunctive particles indicating causality)" [2]. Meanwhile, Sakaji [7] researched a method of acquiring knowledge about causal relations using causal markers for all cause and effect knowledge. Based on these related works, we propose effective causal markers as the measure for distinguishing If-Then delimitation based on the result of analyzing design knowledge information in press releases.

4. VERIFICATION

4.1 Verification of Automatic Tagging Method of Design Tags and Filtering Method

In our experiment, we verify the proposed method by applying it to 20 press release articles acquired by Webstemmer. First, we mark the paragraphs including useful sentences for design in each acquired press release. In this experiment, we use press releases for a time clock, a wool coat, a game controller, etc. We assume these marked paragraphs to be the correct answers. We apply the proposed method to the articles and examine whether the method can output the correct answers. In this research, we assume that several kinds of design tags are put to use in finding information for design. Because of this assumption, if the number of kinds of design tags in a sentence is greater than a threshold value, we regard the paragraph which includes such a sentence as useful information for design. We define the threshold value as the design tag's threshold. We classify the design tag acquisition threshold into four stages with and without the filtering method as shown in experiment 1-8 in Table 3. The method is evaluated according to F-measure [8]. When we compare the F-measure of each experiment without
Table 3. Results of the verification of the automatic tagging method for design tags and the filtering system

<table>
<thead>
<tr>
<th>Design tag's threshold</th>
<th>Filtering</th>
<th>Average accuracy</th>
<th>Average recall ratio</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>One kind of design tag</td>
<td>Executed</td>
<td>9.15%</td>
<td>97.14%</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>Two kinds of design tags</td>
<td>Executed</td>
<td>6.25%</td>
<td>60.00%</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>Three kinds of design tags</td>
<td>Non-executed</td>
<td>14.95%</td>
<td>94.43%</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>Four kinds of design tags</td>
<td>Executed</td>
<td>19.00%</td>
<td>64.95%</td>
</tr>
<tr>
<td>Experiment 5</td>
<td>Five kinds of design tags</td>
<td>Non-executed</td>
<td>22.22%</td>
<td>87.46%</td>
</tr>
<tr>
<td>Experiment 6</td>
<td>Six kinds of design tags</td>
<td>Executed</td>
<td>33.33%</td>
<td>48.57%</td>
</tr>
</tbody>
</table>

The filtering method, we confirm that the F-measure improves proportionally as the number of kinds of design tag increases. Therefore, the automatic tagging method for design tags can result in the efficient acquisition of useful information for designs by setting the design tag's threshold to be high, and we confirm that the method tends to tag multiple design tags as representing useful information for design. The experimental results show the effectiveness of the information extraction method based on the linguistic features of design information. Moreover, as a result of applying filtering system to automatically tagged sentences, the accuracy and the F-measure improve the design tag's threshold in most cases. We can confirm that the filtering system can reduce the noise in the Web information database, and the effectiveness of the proposed method has been demonstrated. In addition, we carry out experiment 9 in order to examine adding the condition to the filtering system based on the result of automatic tagging system of design tags. In experiment 9, design tag's acquisition threshold is three kinds of design tag, and the filtering system is applied to only the sentences put three kinds of design tag. As a result, F-measure improves in comparison with experiment 6. Therefore, we consider that we can improve the proposed method by combining the result of the automatic tagging method for design tags and the filtering method.

4.2 Verification of the Automatic Conversion Method for Knowledge Representation

First, we verified the effectiveness of the method to divide design knowledge into one design knowledge in one line. When we applied this proposed method to 100 press releases, it was able to automatically divide the design knowledge from all of the press releases into one knowledge in one line. As mentioned above, the effectiveness of the method is shown. Next, we verify the effectiveness of the improved method of the automatic document tagging system. As a validation methodology, we manually obtain the correct data for the delimitation of If-Parts and Then-Parts for 32 kinds of design knowledge information described the causal relation (hereafter, we call this "If-Then type knowledge information"), and verified whether the method could correctly process them. The result is shown in Table 4. As mentioned in Table 4, the effectiveness of this method is shown. Meanwhile, this method could not process the If-Then type knowledge information described without causal markers, such as itemization, etc.

Table 4. The result of the verification of the improved method of the automatic document tagging system

<table>
<thead>
<tr>
<th>Design tag's threshold</th>
<th>Filtering</th>
<th>All number</th>
<th>Accuracy before it</th>
<th>Accuracy after it</th>
</tr>
</thead>
<tbody>
<tr>
<td>If-Then type knowledge information</td>
<td>32</td>
<td>85.00%</td>
<td>65.00%</td>
<td></td>
</tr>
</tbody>
</table>

5. CONCLUSION

In this thesis, we proposed a Web-based knowledge database construction method for UAS. We demonstrated the effectiveness of the method. We will improve it based on the findings, and implement it completely in the future.

6. REFERENCES


