An Idea of Formalizing Web page Information Architecture

KISHIMOTO Yorinori
Tokyo University of Information Sciences
Department of Information Systems
4-1 Onaridai, Wakaba-ku, Chiba, 265-8501
Japan

Abstract: For analyzing performance of Web site, a point of view of the information architecture has been proposed.[1] It considers from the structural side in information elements composing Web site, and this idea derives information structure of a high usefulness for a user of Web site. However, it is required that experience of many Web designs and interface designs because this idea includes a difficult concept to use it in a Web design work. In addition, it is difficult to use this idea in a practical field, because to express clearly the Web information architecture is few. This paper proposed a formalization method of Web page’s information architecture by regular expression for checking its structure. This method classifies structure elements of a Web page in attributes on the basis of the Web information architecture, and its expressed by two types of equations on the basis of the F-Shaped reading pattern. This method can verify structure of the information architecture of a Web page. As a result, this method was able to analyze to structure elements of lack point and a redundant point in Web page information architecture.

Key–Words: Web Information Architecture, Formalization, Web design, Interface design

1 Introduction

For analyzing performance of Web site, a point of view of the information architecture has been proposed.[1] It considers from the structural side in information elements composing Web site, and this idea derives information structure of a high usefulness for a user of Web site. However, it is required that experience of many Web designs and interface designs because this idea includes a difficult concept to use it in a Web design work. In addition, it is difficult to use this idea in a practical field, because to express clearly the Web information architecture is few.

Thus, we proposed a formalizing method for express information architecture of a Web page by regular expression. According to the idea of information architecture, structure elements of a Web page can be classified roughly into the browsing support, the search system, the contents task and other invisible elements. In addition, it had been proposed that reading pattern of Web site becomes F-shaped.[2] Then we focused on information architecture elements and reading pattern of Web site users. It is required that the information architecture of a Web page is able to browse a user without confusion and contradiction. In other words, information architecture quality of a Web page can be argued about if based on this structure.

In this method, this structure is expressed by regular expression. At first, information architecture construction elements are abstracted as a variable based on an attribute of the Web information architecture. In this abstraction, a meaning of sentences and/or design elements these are included in construction elements of a Web page can be ignored, and only a meaning as a structure element can be extracted. It can be expressed as a deterministic finite-state automaton by these abstracted elements concatenated along an F-shaped reading pattern of Web site user.

This paper proposed that a method to formalize the information architecture in a Web page by regular expression. In addition, this method was applied on a practical Web page, and it was confirmed that method could apply this example. As a result, whole information architecture of Web page could be expressed by regular expression, and it has been detected to a lack point and redundant point in this example’s information architecture.

2 An Idea

2.1 Basic Idea

In this paper, concatenation of construction elements of Web page is defined as the information architecture of a Web page. It is necessary to considering to naming rules of links and/or a category of classification
system by an idea of the Web information architecture. However, analysis of the Web information architecture becomes difficult because these attributes are mixed for one element.

Thus, we focus on a concatenation of information architecture elements as a purpose to usefulness measurement of a Web page. If information architecture of Web page is poor quality, a user of this Web site feels difficulty. Especially, in a Web page designed based on a definite rule, disarray of context becomes it with a factor harming usefulness. In addition, similar Web page elements are set to be adjacent. For example, a category list of similar items is adjacent in a design of shopping site. Naturally usefulness becomes poor when these elements scattered. In other words quality of a usefulness of a Web page can be discussed on by focusing its attention on the concatenate of structure elements of a Web page.

2.2 A classification of structure elements In the Web information architecture

A classification of Web site structure element is proposed to Web information architecture. On the basis of this idea, information architecture elements were classified. A structure element of a Web page can disregard an invisible element such as thesaurus. In addition, it can be disregarded because query languages in a search support and search algorithm are not structure elements of a Web page element directly. Table 1 shows elements of Web information architecture that derived from this idea. In addition, it was shown in table 1 about an abstraction variable name to show in 2.3. In addition, Table 1 shows abstraction variable names that appeared in 2.3.

2.3 Formalization of the information architecture of Web page

Web information architecture elements are abstracted by variables shown by Table 1. Furthermore, it was abstracted by regular expression along a flow on the F-shaped reading pattern. A structure element of a Web page has a design, a color, a sentence, shape, and an attribute as a structure element of the information architecture. However, this mixture attribute makes difficult to analysis of Web information architecture.

Thus only an attribute of the information architecture is extracted in a structure element, and an extracted element is abstracted as a variable of regular expression. And a regular expression is derived from concatenation of elements along the F-shaped reading pattern.

This method can be shown clearly to the information architecture in a Web page.

| Table 1 Elements of Web information architecture |
|------------------------|-----------------|-------------------|
| Category               | Attribute       | Variable          |
| Browsing support       | Global Navigation| $N_g$             |
|                       | Local navigation | $N_l$             |
|                       | Context navigation| $N_c$            |
|                       | Site map         | $N_m$             |
|                       | Site index       | $N_i$             |
|                       | Site guide       | $N_u$             |
|                       | Other navigation | $N$               |
| Support of search      | Search interface | $S_i$             |
|                       | Search zone      | $S_z$             |
| Contents task          | Headline         | $C_e$             |
|                       | Embedded link    | $C_l$             |
|                       | Embedded meta data| $C_m$         |
|                       | Chunk            | $C$               |
|                       | List             | $C_l$             |
|                       | Continuous step  | $C_s$             |
|                       | Identifier       | $C_i$             |

2.4 An idea of analysis

In information architecture analysis of a Web page, it uses two equation of regular expression that derived from 2.3. One expression treats each structure element as an individual variable. In this case, navigations to have a same link and/or same information are handled as the same variable. By this, a structure of the whole Web page can be expressed. Another expression is a regular expression that was provided by abstracting a variable of a structure element only by an attribute of own. In this expression, concatenation of structure elements is expressed by Kleene enclosure iteration (* symbol). This analysis method uses these both equations.

For example, it considers following two expressions.

$$A = CcC_1C_1N_1C_cC_2C_2N_1C_cC_3C_3$$
$$N_1C_cC_4C_4N_1C_cC_5C_5N_1l_1$$

$$A' = (CcCN)^*CcCNl$$

Equation $A$ and $A'$ are the results that formalized the same Web page, and a meaning of a used variable name is based on table 1. Equation $A$ abstracted structure elements of a Web page to an attribute and an identifier. Equation $A'$ abstracted equation $A$ only in an attribute and expressed an iteration by Kleene enclosure.

Structure of equation $A'$ consists of two parts of $(CcCN)^*$ and $CcCNl$. However, $(CcCN)^*N_1l$ can be in this structure naturally because it can be con-


3 A Formalization Method

An analysis procedure by this method is shown as follows.

Step 1: Structure elements separation

Web page is divided into structure elements.

Step 2: Setting of variables

Variable names appropriate to each attribute in Table 1 and subscripts for its identification are added as the same variable name and subscript.

Step 3: Formalization

A variable provided in Step 2 is arranged along F-shaped reading pattern. In some Web site, there is the case that does not assume F-shaped reading pattern. In this case, structure elements suppose a reading pattern of a user, and it is enumerated along the flow.

Step 4: Extraction of regular expression for analysis

Arrangement of a word provided in Step 3 is abstracted only by an attribute. As a procedure, a subscript is disregarded, and arrangement of elements extracted only in own attribute of structure element is expressed with iteration (Kleene enclosure).

4 An Example

It shows an application example of this method. This example applied it on a Web page of Tokyo University of Information Science[3].

Step 1 & 2: Structure elements separation and setting of variables

Figure 2 shows the division of a structure element in this example and a variable name corresponding to each element.

Step 3: Formalization

An equation to be provided from this example shown as follows.

\[ Ni_1 N_2 N_3 N_4 N_5 N_6 N_T C_1 N_8 N_9 N_{10} S_i_1 S_i_2 \]
\[ N_g_1 N_g_2 N_g_3 N_g_4 N_g_5 N_g_6 N_g_7 N_g_8 \]
\[ N_{l_1} N_{l_2} N_{l_3} N_{l_4} N_{u_1} N_{u_2} N_{u_3} N_{u_4} N_p \]
\[ C_c C_c C_l C_l C_l C_c C_c C_1 N_{11} \]
\[ C_c C_c C_n_{11} C_c C_n_{12} \]
\[ C_c C_n_{15} N_{l_1} N_{l_2} N_{l_3} N_{l_4} N_{l_5} \]
\[ C_c C_n_{16} N_{l_6} N_{l_7} N_{l_8} N_{l_9} N_{l_{10}} N_{l_{11}} N_{l_{12}} N_{l_{13}} \]
\[ N_{t_1} N_{t_2} N_{t_3} N_{t_4} N_{t_5} \]
\[ N_{g_1} N_{g_2} N_{g_3} N_{g_4} N_{g_5} N_{g_6} N_{g_7} N_{g_8} \]
\[ N_{t_{13}} N_{t_{14}} N_{t_{15}} N_{t_{16}} C_4 \] (1)

Step 4: Extraction of regular expression for analysis

It disregards a subscript from an equation provided in extraction Step 3.

\[ N_{i_1} N_{i_2} N_{i_3} N_{i_4} N_{i_5} N_{i_6} N_{i_7} C_c C_n_{i_8} N_{i_9} N_{i_{10}} S_i_1 S_i_2 \]
\[ N_{g_1} N_{g_2} N_{g_3} N_{g_4} N_{g_5} N_{g_6} N_{g_7} N_{g_8} \]
\[ N_{l_1} N_{l_2} N_{l_3} N_{l_4} N_{l_5} N_{l_6} N_{l_7} N_{l_8} N_{l_9} N_{l_{10}} N_{l_{11}} N_{l_{12}} N_{l_{13}} \]
\[ N_{l_{14}} N_{l_{15}} N_{l_{16}} N_{l_{17}} N_{l_{18}} N_{l_{19}} N_{l_{20}} \]
\[ N_{t_1} N_{t_2} N_{t_3} N_{t_4} N_{t_5} \]
\[ N_{g_1} N_{g_2} N_{g_3} N_{g_4} N_{g_5} N_{g_6} N_{g_7} N_{g_8} \]
\[ N_{t_{13}} N_{t_{14}} N_{t_{15}} N_{t_{16}} C_4 \]

It is disregarded in regular expression with iteration in this equation.

\[ Ni(N)^*Cc(N)^*Si(Ng)^*(Ni)^*Np(Cc)^*(Cl)^* \]
\[ (CcCN)^*CcCNl(N)^*Cc(Ni)^* \]
\[ (Nt)^*(Cc(Nt))^*(Ng)^*(Nt)^*Cc \] (2)

In this way, it showed that it could formalize a Web page.
Figure 1: Separation of Elements
5 Considerations

5.1 Consideration of structure element lacking

It shows an analysis of the Web pages information architecture of this example that using equations provided by this method. We focus on the underline part of equation 2. It corresponds to an underline part of equation 1. As for sub equation \((CcCN)^*CcCNl\), As for the sub equation of this, \((CcCN)^*CcCNl\) is more natural than \((CcCN)^*Nl\). Therefore, it is supposed that this structure lacked \(N\) in a designing. In this case this structure should be corrected so that the flow that a user reads by the same shape is to be changed suddenly.

5.2 Analysis for redundant structure

There are two global navigation groups in equation 2. This both structure elements have the same links. Therefore, it is supposed that this part is redundant, and global navigations of the lower part should be cut. In addition, it is supposed that these are needless because navigation \(N_{11}\) returning to the page top just before navigation group is inserted.

6 Conclusion

This paper proposed that a method to formalize the information architecture of a Web page by regular expression. This method expresses the information architecture of a Web page by two kinds of equations formally. One is the equation that keeps structure elements identification. Another is regular expression to have the iteration that it abstracted only in an attribute.

In addition, this paper proposed a method to analyze the information architecture of a Web page from these two equations. As a result of having applied it, this example was able to formalize structure of Web page by this method. Furthermore, by analysis with these two equations, it was able to check a redundant point and lack point about this example’s information architecture.

References:


