

## A study of Efficient Access to homepage

Osamu.Yamagata<sup>†</sup> Masahiro.Sakamoto<sup>†</sup> and Kazuo.Tsuda<sup>††</sup>

<sup>†</sup>Faculty of Information Sciences,Hiroshima City University

151-5,Ozuka,Numata-Cho,Asa-Minami-Ku,Hirosima 731-3194,Japan

Tel:+81-82-830-1666,Fax:+81-82-830-1792

sakamoto@im.hiroshima-cu.ac.jp

<sup>††</sup> Graduate School of Systems Management, The University of Tsukuba

3-29-1 Otsuka,Bunkyo-ku,Tokyo 112-0012 ,Japan

TEL+81-3-3942-6877,FAX+81-3-3942-6829

Tsuda@gssm.otsuka.tsukuba.ac.jp

WWW(World wide Web) is a place where many laymen receive and send the information at the worldwide. A visualization of WWW is very important as one of functions helping to retrieve and browse the information. However, as the information volume displayed at one time is limited by a screen size, it is useful to limit the volume of information capable for users to browse.

In this report, we propose the page-index, which extract its information according to degree of user's interest. This index is a numbers, which relate the user's interest with verbal and structural characteristics of homepages. The homage-index constructed with sequence of this index are useful to help a user's browsing. We constructed the estimating system about an article of commerce and really tried to estimate. From this trial, we verified the usefulness of the page-index, homepage-index.

### 1. Introduction

in chap.4.

Recently, we can easily obtain the various information by means of conversation process on the Web (World Wide Web) by which we are able to browse the massive information related with each others. However the basic structure of the Web is the texts linked mutually, so-called hypertext, and owing to its non-linearity, there are some difficulties such as "disorientation problem" and "cognitive overhead problem"<sup>[1]</sup>.

In this report, we propose an "interest index" which is a function of the verbal and structural characteristics of Web documents and further using this index construct a model which estimates user's interest.

In this report, we discuss a problem of visualization of web-space in chap.2, propose an interest index of homepage in chct-p.3 and verify a validity of this index and homepage index

### 2. A visualization of Web

The basic structure of Web is nonlinear so that there are many routes to obtain the information. One can choose freely among them, but he must select only one. So the "disorientation problem" and "cognitive overhead problem" are taken over.

The research of 3D visualization for large scale database is going on from the view point of visualization of information and "the slowdown of display speed" and "increase of recognition load" are disclosed. Then, it is pointed out that the important subject is not to display a lot of information but to display only required information (decreasing of quantity of displayed one)<sup>[2]</sup>.

The "Natto view"<sup>[3]</sup> and "HyperSpace"<sup>[4]</sup> is an example of

concrete researches for these problems and offers the gentle display process on the basis of "lifting up operation" of the external information (structure of link and characteristics of web page). However, this process is not enough to resolve "disorientation " and "cognition overhead problems" due to neglect of the information and structure of documents.

### 3. Proposal of page-index

The page-index is defined according to user's interests. This index is a property of information extracted from a homepage which is a structured information. Next, we define the homepage-index composed with the page-index and discuss a capability to help user's browsing .

#### 3.1 Construction of page-index

The page-index proposed here is a number which represents a degree of user's sufficiency. The number is calculated with a frequency of keywords which define the user's interest and total words constructing the text. The calculation is division of frequency of keywords by total words and multiplication, with number of link.

The page-index which indicates the interest of users' is defined by following equation(1).

$$px = fa / fi * nl \quad \text{equation(1)}$$

px : page-index

fa : frequency of key word + 1

fi : sum of frequency of all words on page

nl : number of link on page

#### 3.2 Construction of homepage-index

The homepage-index which indicates the interest of users' keeping sequence browsing pages is defined by following equation.(2).

$$hx = \{px1, px2, \dots, pxn\} \quad \text{equation(2)}$$

hx:homepage-index

pxi:ith page-index

i:number of access order

#### 3.3 Support of user's browsing process by home-page index

The page-index has an ability to have a peak so that the frequency of keyword has a peak in a paragraph. If the page-index show a peak, the approximation shape of homepage-index is upward convex and this shape afford the user that the variation of interest may exhale or converge. So the users can effectively browse along the convex form shown by the homepage-index.

## 4 Verification of efficiency of page index, homepage index

We constructed an estimation system and applied this system pages of several home electric products.

#### 4.1 Estimation system

The system is constructed with the following programs.

(1) Keyword extracting program : sum of frequency of keywords in a page text

(2) Word extraction program : sum of frequency of each word in a page text

(3) Distribution analyzing program of interest : Analysis of access order of user's page (HTML-file) and distribution of user's interest from Web-access-log

(4) Table calculation program (EXCEL) : Calculation of Indices

(5) Equation treating program (Mathematica) : Approximation and visualization of homepage-index and distribution of user's interest

#### 4.2 Estimation about home electric products

Table 1 page-index of C-company's wide TV

order	name of page	frequency of keywords	number of words	frequency of words	page-index
1	www.ccc.co.jp	0	1	183	1
2	psindex	2	1	281	2
3	index	5	3	46	52
4	size	5	9	18	333
5	cz11	9	7	78	103
6	k1001	14	1	176	17

Page-index and homepage-index of 2 products are taken from 10 homepages of 5 Japanese corporations and distribution of user's interest is analyzed with Web-access-log of 6 products of one corporation.

(1) Page-index

The user's interests are defined as the following keywords. The key words for refrigerator "thawing(解凍),refrigerate(冷凍),optimum cooling(適冷), freezing cold(凍結), chilling(冷却), and cold storage(冷蔵)". Those for Wide TV are "image(画像), width(奥行), size(外形寸法), image quality(画質), high quality circuit(高品質回路), electric power(消費電力), digital(デジタル), dome speaker(ドームスピーカ), and wide TV(ワイドTV)". The page-indices are estimated with the homepage of A-, B-, C- D- and E-companys. As an example, the page-indices of C-company's wide TV are shown in Table1.

(2) Homepage-index

The homepage-indices of refrigerators are estimated with the homepage of A-, B-, C- D- and E-companys.

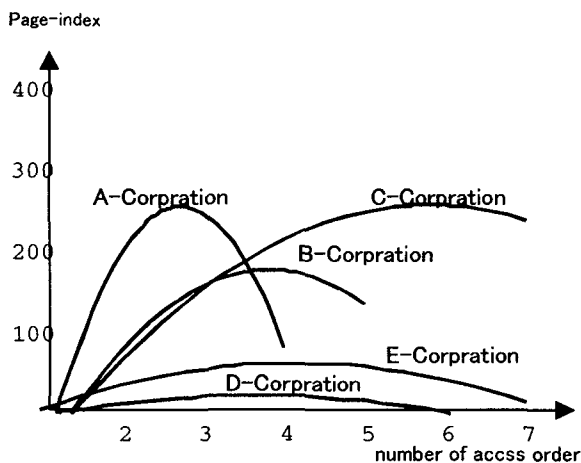


Fig.1 Homepage-indices(refrigerator)

The ordinate of Fig.1 is homepage-index and abscissa is layer number. The lines indicate the approximate state of homepage-indices.

The homepage indices of "Wide TV" are estimated with the homepage of A-, B-, C- D- and E-companys. The ordinate of Fig.2 is homepage-index and abscissa is layer number. The lines indicate the approximate state of homepage-indices.

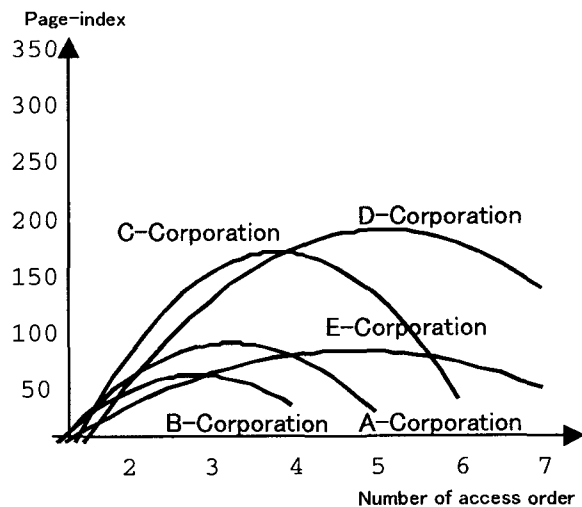


Fig.2 Homepage-indices(Wide TV)

(3) Distribution of user's interest

Fig.3 shows a distribution of user's interest in 6 products of one corporation. The ordinate is number of users and abscissa number of browsed pages. The line indicate the approximate state of distribution.

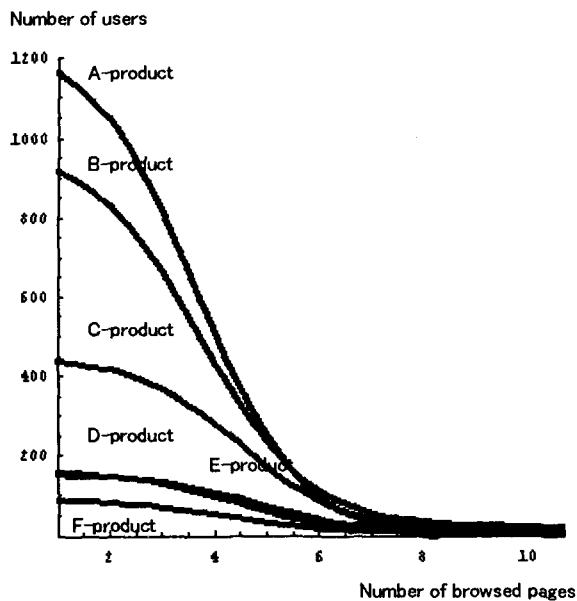


Fig.3 Distribution of user's interest

#### 4.3 Estimation about home-electricity products

As an example, these indices were calculated from the Web-documents of five manufacturers of Japan and the distribution of user's interests from the Web-access-log (monthly log of 6 products of one manufacturer). These results are shown with visualization by Mathematica in Fig.1(indices of 5 homepages of refrigerator) and Fig.2(indices of 5 homepages of Wide-TV), Fig.3(distribution of user's interests in 6 products). It is seen clearly from Fig.1 and Fig.2 that indices show upward convex form. In Fig.3, it is seen that the user leaves rapidly from the homepage at a value of browsed pages, at the rapid decreasing point of interest. This state is a form of differentiation of homepage-index and indicates that this number is founded on the degree of user's interest. From this analysis, the point at which the user's interest changes rapidly is obtained. From these facts, the proposed

page-index is considered as a equation which affords an end or close.

As mentioned above, it is verified that the page-index is effective to a new subject of visualization as "a reduction of displayed information".

## 5. Conclusion

In this paper, we propose the page-index to solve the intrinsic problem of Web and then it is shown that this index is very effective to decide whether the user continues his browsing or not. The page-index proposed in this work, is one method to solve "disorientation problems and cognitive overhead problems" by visualization.

#### Acknowledgement

This work was performed by encouragement of Pro. Morisue.

#### Reference

- [1] Conklin, J. "Hypertext: An Introduction and survey" IEEE Computer vol. 20 no. 9 P17-P41 1987
- [2] Koike Hideki: "Interactive 3-dimension Information Visualization", Computer Software, vol.11, no.6, P20-P31 Nov. ,1994
- [3] Nattoview : <http://www.myo.inst.keio.ac.jp/grrroups/IPS/Nattoview>
- [4] HyperSpace: <http://www.cs.bham.ac.uk/~amw/hyperspace/www95/>