Ubiquitous Networking for GENES Society: e-Learning Tools and Digital Archives for Education with Significant Use of Cultural Heritage Contents

Susumu Sawai

Gakujoken
23,NangenCho,ShinjikuKu,
Tokyo,160-0012,Japan
NAC02440@nifty.com

Toshiko Wakaki

Shibaura Institute of Technology
307 Fukasaku,Saitama-City,
Saitama, 330-8570 Japan
twakaki@sic.shibaura-it.ac.jp

Mitsuru Ikeda

School of Knowledge Science, JAIST
1-1,Asahiadi,Tatsunokuchi,Nomi
Ishikawa, 923-1292, Japan
ikeda@jaist.ac.jp

Susumu Kunifuji

School of Knowledge Science, JAIST

1-1,Asahiadi,Tatsunokuchi,Nomi
Ishikawa, 923-1292, Japan

kuni@jaist.ac.jp

ABSTRACT

Ubiquitous-networking is an active e-learning environment using the information communication technology (ICT) like cellular phones, personal computers, PDA and so on, where a learner can learn anytime, anywhere and any style. In this paper, We emphasize here that learners, and especially children should be able to use digitalized cultural heritage for education, as a "digital Cultural Genes". In order to backup "e-Japan strategy", we propose new creation tool

Categories and Subject Descriptors

K.3.1 [Computer Uses in Education] Computer-Managed Instruction

General Terms

Algorithms, Management, Human Factors, Standardization, Language.

Keywords

Digital archives, cultural heritage, ubiquitous networking, digital cultural genes, LOM, RDF, Semantic Web, XML, digital archivist

1. INTRODUCTION

Ubiquitous networking is an active e-learning environment using the information communication technology (ICT) like cellular phones, personal computers, PDA, electronic tag and so on, where a learner can learn anytime, anywhere and any style.

The Ministry of Education, Science, Sports and Culture (MEXT) founded National Information Center for Educational Resources (NICER) in August 2002.

NICER has the roles to arrange and manage all the information about education and learning in Japan on the Internet. In this paper,

we propose new creation tool in order to backup e-Japan strategy.

Fig. 1 shows the relation between GENES (Gakujoken Network Studying group) and NICER.
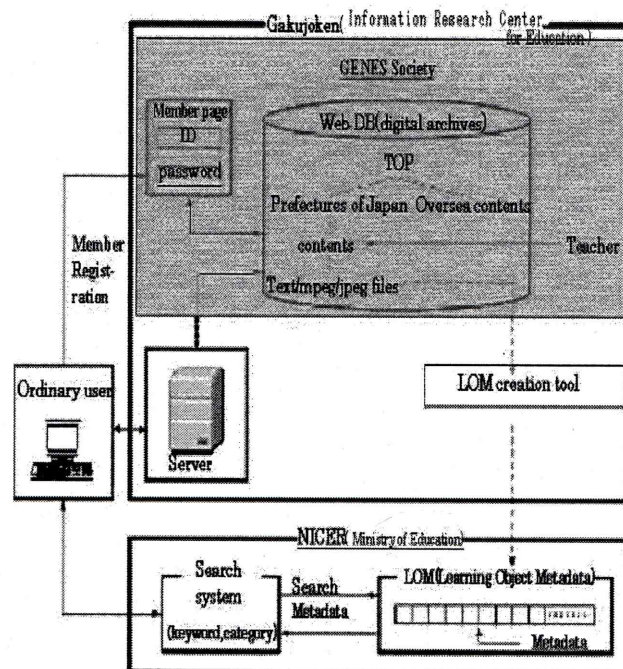

Fig.1 The relation between GENES and NICER

In this decade, authors have been coordinator of several Japanese Government ICT projects for policy preparation under the

Minister of Economy, Trade and Industry (METI) responsibility: Multimedia, Digital Archives, applications of ICT in Education.

GENES Project was launched in August 2001, in cooperation with Fujitsu and Ministry of Education[1] . GENES stands for Gakujoken network studying group. Gakujoken is the public organization managed by MEXT. The roles of Gakujoken is to collect and to offer the educational contents to learner.

2. OBJECTIVE OF GENES

The objective of GENES is to create the society where cultural heritage contents are created by teachers, as the knowledge base of digital archives for education. The roles and aims of GENES are to realize the e-bok like electric book that learners, especially children should be able to use digitalized cultural heritage for education, as a "Digital Cultural Genes". Today the profile of the project to realize GENES society is as follows:

-Cooperation with more than 1200 school teachers.

-More than 4,500 learning object metadata(LOM) for educational contents.

-More than 37,000 educational contents (motion picture, photos and texts) are combined in this digital archives. These educational contents contain the following specifications: a school grade, a subject and a unit, a coverage person, a coverage date, a coverage place, copyright, etc.

In the next step, we want to use the concept of Semantic Web, and digital archivist. We try to build the LOM RDF binding and defining LOM metadata by simple methods of knowledge representation, the Resource Description Framework, RDF[2].

Digital archivists are the professionals supporting the cultural activities in an information society. A digital archivist digitizes the cultural data which are needed in the future fields. Simultaneously, a digital archivist understands copyright and privacy, cultural art, protection and management, and so forth as the foundation of cultural activities.

3. KNOWLEDGE REPRESENTATION ON PAIRS OF "TACIT KNOWLEDGE" AND "EXPLICIT KNOWLEDGE"

Tacit knowledge[3] established through still/ motion pictures is a more suited approach for intuitively storing information to the human memory compared to explicit knowledge acquired through literary tools. Therefore it can be said that tacit knowledge is an optimum method for understanding the context (status) of oneself or another.

Advantages of implementing the new knowledge representation based on the combination of motion pictures and metadata are that it enables text-search within the realm of tacit knowledge derived from still/motion pictures.
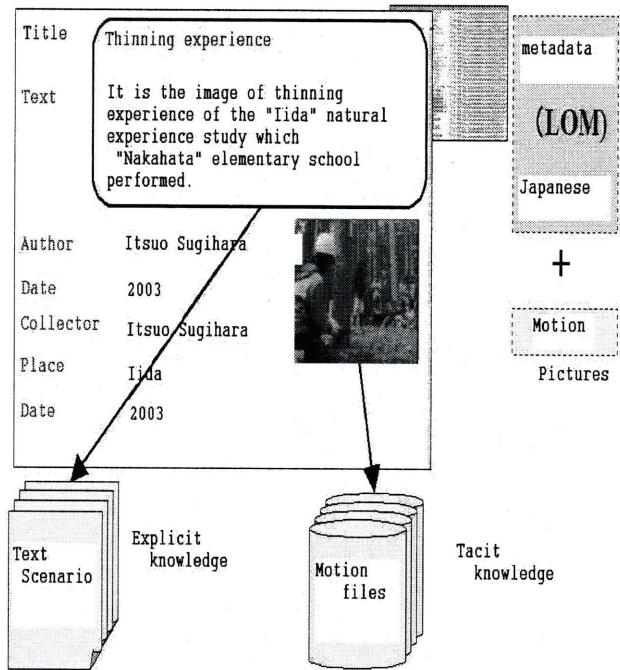

Fig.2 Knowledge representation on Pairs of "Tacit Knowledge" and "Explicit Knowledge"

3.1. Applying tacit knowledge to Web page Search Systems

Web pages targeted in this research consist of text scenarios of explicit knowledge and still/ motion digital image files of tacit knowledge. The explanation of the motion file is in the text scenario database.

The benefits of this are that sharing methods utilizing web-page search systems will become available, for example the application of tacit knowledge through motion pictures to e-learning.
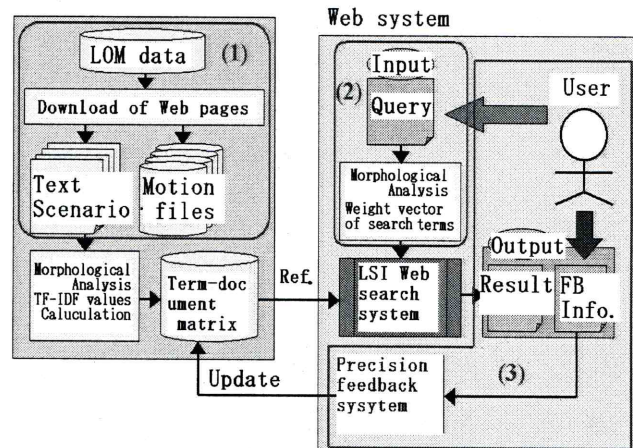

Fig.3 Web page Search Systems

Web pages similar to the imputed query will be searched within the database created by the text scenario of web pages, using Latent Semantic Indexing (LSI) [4][5] as a search method. Learning Object Metadata (LOM) which enables search for digital archives will serve as an interface for the actual search. This will enable the application of National Information Center for Educational Resources(NICER) to the search system.

In addition, this will help applying tacit knowledge from motion files to e-learning as well as conducting feedback searches.

Search criteria should be the sentences close to natural language. This will enhance universality and contribute to raising the system's utility value.

3.2. Establishing Term-Document Matrix

While necessary information such as ID numbers, titles, URLs will be extracted from the LOM data and stored as a file, index terms will be extracted from the text scenario downloaded from the targeted URL. The extraction will be conducted by morphological analysis and TF-IDF value calculations. Subsequently, an index term - document matrix will be established.

3.2.1 Morphological Analysis

Assuming that nouns well represent the characteristics of a web page, the target text will be morphologically analyzed in order to extract only the nouns included in it. Morphological analysis will be conducted using a Japanese language morphological analysis tool "chasen"[9]. Numeric characters/ pronouns will be deleted as stop list.

3.2.2 TF-IDF Value Calculation

The weighted value of each noun (TF-IDF) will be obtained based on calculation of the value indicating the frequency of its appearance in each page (TF) and the value indicating the frequency of its appearance in all documents (IDF) . Nouns with a value exceeding a certain threshold value will be extracted as index terms.

3.2.3 Structuring the index term – document matrix

Each web page will be described as a document vector in the vector space dimensioned by the extracted search term, and its element will be the TF-IDF value. The search term – document matrix is a matrix representation of such document vector sets.

3.3　Web Search system by LSI Method

A web page with high similarity will be searched/ output by generating a weight vector of a search term in response to a query sentence in natural language and referring to the index term – document matrix using the LSI method.

3.3.1 Latent Semantic Indexing (LSI)

Latent Semantic Indexing will search base matrixes with ranks lower than the original index term – document matrix by conducting singular value decomposition against index term – document matrix and condensing dimensions with small singular values. This is a method to condense document vector dimensions by projecting each vector onto this base.

When an index term – document matrix is given as $m \times n$ matrix D, the singular value decomposition of D can be defined as follows:

$$D = U \Sigma V^T$$

U is a $m \times m$ orthogonal matrix ($UU^T = U^T U = I$), V is a $n \times n$ orthogonal matrix ($VV^T = V^T V = I$), Σ is a $m \times n$ matrix. When $rank(D) = r$, the number of singular values σ_i diagonally arranged in descending order will be r.

The projection onto U_k space can be considered in order to approximate a r -dimensioned document vector d to a k - dimensioned document vector $d^{(k)}$.

U_k is a $m \times k$ matrix composed only by the first k ($k < r$) left singular vectors of U.

$$d^{(k)} = U_k^T d$$

The index term – document matrix will be condensed to a lower dimension in this form.

The similarity is found by taking the cosine $sim(d,q)$ against a document vector d and query vector q in a vector space. The similarity $sim(d,q)$ can be calculated from U_k, Σ_k, V_k without seeking the index term – document matrix condensed to k dimension(D_k).

3.4　Precision feedback system

Users count the right answers from search results, and input this correct information to the Web search system. Then search ability of this system should be improved by Precision feedback system.

3.5　Evaluation of the Search System

The assessment of this Web search system will start by imputing natural language query and the number of dimensions/ results to display. The server will receive such data, calculate the similarity and display the search results.

The search results will display in descending order of similarity against the imputed data as shown in Fig.4.

It can be seen in Fig.4 that keyword search against the same evaluation subjects could not retrieve corresponding web pages.

Table 1 indicates that the search results are reasonable. However, as the current data volume 136 documents is not sufficient, it is desirable that the assessment of search precision be considered upon increased amount of data.

Natural language query

What is the history which Mr. Nobunaga Oda went to a battle of "Okehazama" war?

Search results

ID	Title	Similarity
110000000055	"Atsuta" Shinto shrine.	0.478389
110000000059	The place said to have shot Mr. Yoshimoto Imagawa.	0.411242
110000000060	The "Nobunaga" wall in "Atsuta" Shinto shrine.	0.360515
110000000058	The place where Mr. Yoshimoto Imagawa established headquarters.	0.360025
110000000053	The castle tower of the present "Kiyosu" castle.	0.336507
110000000056	The army corps course of a battle of "Okehazama" war.	0.290821
110000000144	The situation of "Washizu" fort.	0.245510
110000000057	The ancient battlefield of "Okehazama" war.	0.204634
110000000140	The situation of "Otaka" fort.	0.201269
110000000065	The capture figure of "Nagashima" riot.	0.183833

Fig.4 Search results of Web page Search Systems

Right	136Dim.	100Dim.	50Dim.
Query1	19	18	17
Query2	14	14	12
Query3	14	15	14
Query4	12	12	12
Query5	17	18	18
Average	15.2	15.6	14.6

Table1 Evaluation of search results

4. THE BULLETIN BOARD FOR UBIQUITOUS-NETWORKING CORRESPONDING MULTI-CAREER

The description method of the contents on the Mobile computing varies from 3 types of cellular phones, i-mode, EZWeb and J-SKY in Japan. As seen from a multi-career correspondence view, there are the following problems:

==

- Inefficiency of the individual contents creation for 3 types of cellular phones

- The extendibility to a new standard is low.

- URLs accessed by 3 types of cellular phones are different.

The anywhere anytime paradigm are realized by the extensible Markup Language (XML). First of all, all of the data written in the bulletin board and the data which can be displayed are saved in XML form. Then the screen data of the bulletin board created by XML are transformed automatically, using XSLT.

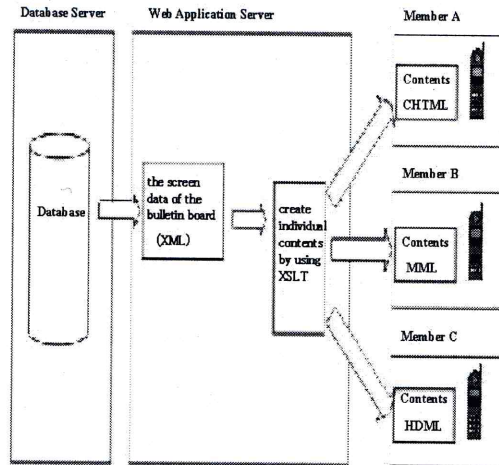Fig.5 shows the automatic content transformation flows, corresponding to each cellular phone.

Fig. 5 Automatic content transformation flows

An administrator does not need to create individual contents for each cellular phones, and creation and management of contents become easy , as the Mobile computing of one-source multi-use styles.

Moreover, by adding XSLT, a correspondence new model can be easily added. It is easily extensible. Any user can access the same URL from any types of cellular phones, i-mode, J-SKY, or EZWeb.

5. CONCLUSION

We emphasize that learner, especially children should be able to use every digital Cultural Heritage for education, as a "Digital Cultural Genes". In This paper, we propose new creation tool in order to backup "e-Japan" strategy.

Anywhere anytime paradigm of Ubiquitous networking becomes more popular for learners to utilize the knowledge-base of digital archives for education made by GENES society.

5. REFERENCES

1) Susumu Sawai, Tosiko Wakaki,Tomio Saku,Junnichi Yamanisi: Mobile and Ubiquitous Networking for GENES Society, WWW2004,New York,2004

2) Nobuo Saito, Tasuya Hagino : Semantic Web,Ohom,2004

3) Michael Polanyi : THE TACIT DIMENSION, Routledge & Kegan Paul Ltd., London (1966)

4) Kenji Kita, Kazuhiko Tsuda, Masamiki Sisibori: Information Retrieval Algorithms,Kyouritsu,2002

5) Tokunaga, Higuchi, Wakaki, Nitta : Development and evaluation of Similar Document search system from judicial precedent DB,2004