2000年度研究報告会抄録

Researchers of Information Sciences in Japan and their Research Fields

Masaki NISHIZAWA, Yuan SUN, Masaharu YANO

The research fields of informatics or computer-related sciences are very wide. It is said that Japan is far behind the United States in these fields, but there are not many persuasive data for it. In this paper, we give some statistical comparisons between Japan and the United States about the number of researchers, the number of papers per fields, and so on. They are done using "Directory of Researchers" Database by National Institute of Informatics, and so on. Further, in order to investigate the content of information sciences and the validity of the classification of "Directory of Researchers" Database, we compare the declared research fields and the fields of their research themes with the classification by "Handbook for Information Processing" by Information Processing Society of Japan. Also, we analyzed the relation of three research fields declared by each researcher.

Tendency analysis of article citations in Japanese academic disciplines

Yasuhiro YAMASHITA, Masaki NISHIZAWA, Yuan SUN, Masamitsu NEGISHI

In recent years, research evaluation is becoming one of the major topics of governmental science policy. Along with this trend, citation analyses attract great attention as a tool for quantitative measure of quality of research output. We report some results of an analysis on the characteristics of citations in academic disciplines, based on the ISI citation statistic database: National Citation Report (NCR). This database contains bibliographies of articles written by authors affiliating to Japanese institutes from January 1981 to June 1997, with yearly citation counts.

First, we analyze frequency distributions of citations by academic disciplines, and then investigate changes of citation counts per article through the years after the publication. Finally, we classify all disciplines into four types by the two indices: number of papers and cited times per article.

Some Trial for Making an Expanded Paradigm of Information and Knowledge Science

Shigezo Murakami

Abstract: We have new images of paradigm based on Japanese culture and Asia civilization. Upon these paradigm we are making the tentative specification of "Data Complex" which is an expanded concept for Data Base.

Extraction of Hierarchical and Associative Relationships among Terms in Consideration of Exceptions

Takayuki Morimoto, Yuzuru Fujiwara

Information technologies are being developed at unprecedented speed due to high performance and inexpensive computers and Internet have been widely available. The transmission and utilization of information become more diversified and borderless very rapidly. However, users may not make good use of huge amount of information by using conventional computers whose major functions are numerical calculation, symbol matching in information retrieval and deduction. Therefore, advanced utilization of contents of information is required gradually.

Learning and thinking are worth a while targets to such requirement and have been widely studied without useful results thus far. In order to realize machine learning and thinking, it is necessary to know meanings and characteristics of terms and various relationships among them, because technical terms are the most convenient and powerful representation medium of abstract concepts. Therefore, the methods of constructing organized knowledge resources are based on extracting semantic relationships among terms. However, there are exceptional terms which may not be bypassed in natural languages. In this paper, we report the method of extracting hierarchical and associative relationships including such exceptions.

New approach to the Wuster's four field term model as a model of systematizatio and creation of knowledges

Yutaka Okaya

This paper tries to investigate the Wuster's four field term model from the viewpoint of systematization and creation of knowledges.
1. The four field term model is effective both in systematization of knowledge, for example cont rolling of synonym and also in creation of knowledge, for example metaphor
2. The table of concepts is effective in explaining the network of concepts.
3. Some discussions about concept and further proposal, ie. constructing a creativity assisting system are presented.

A Fundamental Study on the Information Modelling of Multidisciplinary Objects

Jyunki YAEGASHI The Faculty of Information, SHIZUOKA Univ.

Despite a rapid growth of the social interest in information and related technologies, Japan falls far behind countries in constructing the social information infurastructure such as databases. Vast majyority of objects with educational, academic, and/or cultural importance are left unexplored. This study ales to establish a basis for the analysis and modeling of information conveyed in these previously unexplored multidsciplinary objects, and to systematically transform it into a social information resource. We demonstrate the feasibility of systematic modeling of these objects. Our approach here is based on the modeling of information activity as a communication activity, and objects concerned as the media for this activity, thereby revealing the characteristics of interdisciplinary information activities.

An application of XML for the microbial biological information

Satoru MITAZAKI , Kohji TAKAHASHI , Hideaki SUGAWARA

Thanks to the genome projects and the research for the bio-diversity, we can find a great amount of biological information in Web servers. Thus it is very useful for us to set up a common mechanism to retrieve and exchange biological information worldwide. One of the most fundamental parts of the information sharing is to design the extensible format for the target data in diverse domains of biology. We propose a format by use of XML for the description of the microbial biological information such as bio-chemical characteristics, DNA sequences and their annotations.

Search and Management Technology for Large Number of 2D Images

Hironobu Morikawa , Naoki Imai , Akinori Nokami , Takehiko Tanaka , Masaru Nakagawa , Wakayama University

We developed a system which enables users to register a large number of 2D images efficiently and to retrieve the image which matches

his or her wish. In this system, the user can register, modify and delete the image data together with attributes through a Web browser, since CGI programs which the Web server executes access the image database. An experimental result shows that this method reduced the registration time to about 1/7 against the traditional way of user's inputting SQL sentences. For image retrieval, we attempted to construct the category structure according to which the image data are classified. Furthermore we developed experimentally a retrieval system for images of tray (or "Bon") and nests of boxes (or "Jubako"), and verified that the user can retrieve the expected images easily. We also use the function of similar image retrieval to detect the duplicate image registration, and to help the more detailed image retrieval.

Question-Clusters-Based Database Its concept and applications

Schu Hirata

Burst of information in quantity has brought a paradox that people have more difficulties to find out necessary information. For the solution, many studies on the information retrieval have been conducted and not a few tools for searching have been developed. But many have experiences that they could hardly pin-point particular information they required. This paper, however, does not propose a new retrieval technology. We have observed that most of non-specialists, particularly Japanese are not well-aware of their problems they want to solve, more precisely, they can not fully define what they want to know. It will be useful if databases are provided with a structure quite different from the existing ones searched by use of key words. We called the new type database as "question-clusters-based database".

Empirical examination on performance of some statistical methods for Japanese text retrieval by using large test collection

Kazuaki KISHIDA

The paper reports some findings from an empirical study on comparison of retrieval performance between some statistical methods: vector space and probabilistic models. A large Japanese text test collection provided by the NACSIS was used, which consists of about 330,000 records of scientific proceedings. Each statistical method was testified using three kinds of indexing techniques for Japanese text: (1) longest matching against entries in a dictionary, (2) tokenizing by change of kind of characters, (3) a simple bi-gram method. Almost no statistically significant difference among the methods was observed, but it seems that probabilistic method based on logistic regression model indicates relatively better performance than other methods.

Building a Browsing System of Metadata

Tetsuo Sakaguchi . Kyoko Shimada , Kana Numajiri , Koichi Tabata

The Library of the University of Library and Information Science launched its digital library service (ULIS-DL) in February 1999. The major functions of ULIS-DL are to build and provide a collection of metadata of network resources for libraries and library and information science. This paper describes building a browsing system of metadata. The system provides a list of words included in the metadata. Users choose one or two words from the list and get a subset of the metadata which include the words. It also provides a list of the creator element of the metadata and a list of Japanese noun phrase which includes chosen word. Users easily get metadata for accessing network resources with this system. The system is available on the Internet.

J-STAGE: Electronic Journal Publication/Dissemination Center: Total system for electronic journal publication and distribution via the Internet

Yukitaka Matsubara

J-STAGE (Japanese Science Technology Information Aggregator, Electronic) is an integrated system which aim to support submission of manuscripts, peer-reviewing, composition and dissemination of electronic journals. J-STAGE is so designed that is freely accessible by participating academic societies or printers via the Internet, and is available 24 hours/365 days. Electronic data is captured in SGML format. J-STAGE provides definition and converter. All transactions including submitting manuscripts, downloding, them for review, inputting review result, and checking review status are done via the web interface. Citations are rinked to external databases and/or electronic journals. Electronic journal data thus created are browsable and searchable via the Internet. J-STAGE also supports production of abstracts for meeting.

XML-based Electronic Publishing of Chemical Article's Multilingual Full-text Database

Ying li , Hidehiro Ishizuka

In an electronic publishing field, attention is being focused on XML. We constructed chemical article's multilingual (Japanese, Chinese and English) full-text database based on XML, and distributed it in the Web. The On-line Journal XML-DTD of NII (National Institute of Information) is adopted as the DTD of the database. Unicode is used as an encoding character code to represent those languages. To display the database in Microsoft Internet Explorer (IE) 5.0, the data of the standard style sheets for XML (i.e. CSS and XSL) are added to the XML document. Applying VB Script and DOM, the elements selected by a user, such as title, author, abstract, bibliography, are extracted from the full-text database, and displayed in IE 5.0. Finally, we discussed the problems related with Unicode, expression of chemical formula, as well as schema.

The Influence on the Commercial Publishing Business By the Internet

Takushi Fukami (Toppan Printing Co., Ltd.)

In this paper I described The Influence on the Commercial Publishing Business by Internet Technology in Japan. Specifically the business structure , the process of the publishing and the distribution channels in an ever-changing environment. The revolution of business process reengineering has occurred and is developing at a very high speed in most business fields. Then I described the problems of the distribution of digital contents such as the publishing contents via the internet.

There are two main problems with the distribution of digital contents via the internet. One is the display of the information. The other is the payment of fees. As of now we can't get the information from the display unit such as we can from paper. Also the publisher can't receive the fee from the consumer who can get the information from the internet, due to a luck of safety mechanisms. There is now no system to ensure the safety of the payment, nor is it user-friendly.

The Impact of Commercialization on Scholarly Communication

Kotaro Nawa

Understanding for 'authorship' in scholarly communication is inconsistent with that for 'authors' right' in copyright law. The inconsistency has become actual under the commercialization of scholarly publishing.