A Questionnaire Survey on Research fields of Information Science in Japanese Universities
Masaki NISHIZAWA and Yuan SUN

A former type classification is not suitable for present-day research organization. Especially in
the information study, the fields are expanding with including other related domains. Then, the
questionnaire survey was performed to 1800 researchers in the information study about their
research field, contribution paper, etc. Here, mainly we report about the relation between the
actual field condition and the contribution paper by the questionnaire survey.

Measuring the Relationship among Japanese Academic Societies with the Cross Citations Found
in Their Papers, and Its Practicability
Yuan Sun, Masamitsu Negishi

This study is based on the citation data extracted from the Citation Database for Japanese
Papers (CJP), produced by the National Institute of Informatics (NII). Institute of Electronics,
Information & Communications Engineers (IEICE) and other 13 academic societies which have
strong citing or cited relations with each other have been chosen for the analysis. Converting
scheme of the raw citation counts into dissimilarity data, and statistical analyses by multi-
dimensional scaling method to measure relationship among the societies are described. The
potentiality of bibliometric research on the CJP database for practical applications will also be
discussed.

Various Problems on transcription of the Japanese Classical Literature Database.
Kazuyuki YASUNO (National Institute of Japanese Literature)

In this report, various problems arisen in the course of the conversion processes from the full-text
database for Japanese Classic Literatures to XML database will be described. This database
had been described under so-called KOKIN rule, however it is necessary to set up a definite
DTD. Furthermore, it will be required to expand W3C Ruby annotation for the expression of
the complex ruby, which is specific in the classic literatures for the conversion to XML. It is
not easy to express the Kanji used in classic literatures. In this report, we solved this question
by applying “Konjaku Mojikyo”. The remaining problem is how to describe the conversion
points of “Kanbun” and how to express the Kanji, which may not be expressed by “Konjaku
mojikyo”.

- 93 -
The Present State and Problem of XML Publishing
Kosaido, Takushi FUKAMI

The Japanese economy is now in lower growth. The publishing industries and also the printing industries are in minus growth for 6 years. The expectation of cross media of printing, packaging and network media we have, but the business of these is not successful now and only the new technology is hot. In this paper I describe the present state and the problem of XML (eXtensible Markup Language) publishing.

Approach to Problems of Developing a Browsing System of Multilingual Metadata
Tetsuo Sakaguchi, Ikue Hizume, Hiromichi Kato

The Research Center for Knowledge Communities at University of Tsukuba introduced the Knowledge Community Information System (KCIS) in February 2003. KCIS inherited the metadata of the digital library system of the University of Library and Information Science (ULIS-DL). In KCIS, metadata are able to be written in several languages because they are expressed in XML and Unicode. The browsing system of metadata of ULIS-DL were designed to handle metadata written in Japanese. This paper discusses problems of applying the browsing system to multilingual metadata. One of the problems is that the system extracts words from metadata with Japanese morphological analyzer. Another problem is synonyms between different languages. This paper describes a prototype system which are developed to make the issues clear. It also describes development of a experimental system to display multilingual XML documents. It is needed for using the browsing system of metadata with lightweight terminals.

Searching System of Researcher Information for Business-Academia Collaboration
Takehiko TANAKA Takayuki HIRANO Masaru NAKAGAWA

We developed a searching system which takes a company profile as an input and retrieves the information of papers or researchers. For the system to extract the keywords from company profiles, papers and researcher information, we focus on complex noun phrases included in the documents. Actually, each keyword is weighted not only according to how often it appears, but according to how often the constitutive nouns appear. We took, furthermore, experiments for the comparison between our system and Namazu, a searching system. As a result, our system is 0.76 time as much as Namazu with regard to recall, while it is 3.4 times with regard to precision.
9 Situations of intellectual properties related on structure and function of proteins
Nemoto, Tadashi

The intellectual properties on protein structures and their functions have been discussed as a major interest of post-genome sequencing era. From this point of view, international activities in the International Structural Genomics Organization, the Human Proteome Organisation and Trilateral Projects by JPO/EPO/USPTO will be introduced. Current situation of intellectual properties on proteins are overviewed and summarized.

10 Extraction of Relationships of Cause and Effect in Patent Documents
Daisuke ISHIKAWA, Hidehiro ISHIZUKA, Norihiko Uda, Yuzuru FUJIWARA

This work is that tring to extract the relationships of case and effect between method and effect to notice specially essential description of invention’s method and effect in patent documents by text process using syntactic rule. we also extract other relationships based on getting relationships. Then, we mention thinking machine model using extracted relationships cause and effect. Notice this work use distributed patent corpus of NTCIR-3 in 2002.

11 「標題学」方法序説
相田 潤

The introduction for the method of “Studying a Title”  
Mitsuru AIDA

“A title” is not the work itself. But the relation with the work to which it is given is very close. And it not only expresses and explains the contents, but as for it, the charm of a work influences them intuitively. In the present age, a “title” is in various fields. Furthermore, it may induce a large amount of wealth. However, the research which tackles it is rare. Though it was familiar, since the example is too much huge, researchers will not be able to hold a clue. However, the start of research is being opened by development of information machines and equipment and resources. By this report, I will try the consideration about the “title” in connection with the classic family register of Japan and China. I want to specifically try the analysis from viewpoints, such as ontology of succession nature, acceptance, and a category title.
The global flow of information is being developed at unprecedented speed. Advanced utilizations of contents of information are required. In order to realize such sophisticated utilization, it is necessary to understand meaning and characteristics of information. Therefore, the structuralization is required to represent various semantic relationships among information. In order to satisfy such requirement, we proposed a new representation of such structure, and made a system for self-organized knowledge resources based on semantic relationships and an application using conceptual structures. However, this system is a prototype and cannot make enough conceptual structures to realize sophisticated utilization. Semantic relationships among knowledge resources must be correct and appropriate to objectives of applications. The main reason is that advanced utilization consists of navigations based on semantic relationships. This paper reports improvements at the method of an automatic extraction of hierarchical relationships which called SS-KWEIC.

Charm Analysis Support System for “Dao-fa Hui-yuan” Research
Norihiko Uda Fumi Tamezawa Koichi Matsumoto and Yoshihiro Nikaido
This paper describes a charm analysis support system for “Dao-fa Hui-yuan” research. The system consists of a support function of name analysis of charms and a support function of parts analysis of ones. The name analysis support function provides KWIC analysis and N-gram analysis. KWIC analysis support function performs order influence arrangement and reverse influence arrangement with the inputted character sequence as the starting point. N-gram analysis support function displays N-gram strings from the 1st place to the 50th place in order of frequency. Charm parts analysis support function manages the parts information which constitutes a charm.

A Theory and a Computer Tool for Quick Understanding of an Economic Article
Schu Hirata, Makoto Yokoyama
While search technologies have made marked progress in recent years, it is yet time consuming and laborious to read references output after searching. Particularly so, to Japanese reading English journals. From our experiences, we knew that sentences having numbers are more meaningful in an article. We came to an idea that if such number-containing sentences are extracted, it will be helpful for a reader to quickly grasp an outline of the article. This report is a result of our examination of the theory, and we developed a PC tool for extracting such number-containing sentences.
Materials design is a typical inverse problem to find out certain atomic constitution with required properties based on available information. Although in the long history of development of new materials, the experience has been played an important role, the data-driven discovery approach which based on well organized materials data is becoming an powerful tool nowadays with the drastic progress of informative technology. PAULING FILE is a comprehensive database for alloy, intermetallic and inorganic binary materials with containing structure, diffraction, constitution, and physical property data published within the last 100 years. The newly released PAULING FILE, Binaries Edition contains about 28,000 structure entries, 27,000 diffraction entries, 43,000 property data and 8,000 constitution entries and 8,000 images of phase diagram. Searching within this huge amount data from various aspects would reveal the “hidden” regularities and correlation and directly provide hints on candidate materials in preliminary stage.

We improve the algorithm to align amino acid sequences of identical protein which is one of the most fundamental operations studying the analysis of genome. In pair-wise alignment, one chooses one aligned pair (i.e., two sequences) without special reasons from several aligned pairs (the number of these pairs is often very large) giving the same smallest values to the difference properly defined between two sequences. In this paper, we compute the mutual entropy for several such pairs having the same difference, and we classify the pairs into some groups such that the same group consists of the pairs having the same value of the mutual entropy, then we finally compute the mean value of the mutual entropy over the whole groups. As a consequence, we can observe the following interesting fact for some proteins that the aligned pair obtained by usual alignment with geometrical protein structure (we call such a alignment the biological alignment here) is in the group having the value of the mutual entropy closest to the mean value of the mutual entropy. From the above observation we conclude that our method using the alignment (MOU-alignment) and the mutual entropy makes us possible to find the biological alignment, that is, we do not need to know the geometrical structure to obtain the biological alignment.
Quality improvement of tRNA gene annotations in the major DNA databases

Takashi Kunisawa

Transfer RNA gene sequences of plastid genomes are extracted from the Gen-Bank/EMBL/DDBJ databases (ISDN) and are aligned to identify tRNA elements. During the course of this effort, mis-annotations of the coordinate and tRNA-assignment (Ile-tRNA, intitiator Met-tRNA, or elongator Met-tRNA) are found at high frequencies. Further supports for high quality database-annotation are necessary to avoid these mis-annotations.