

Journal of Japan Society of Information and Knowledge

情報知識学会誌

Vol.10 No.2 (Jul. 2000)

~~~~~ 目 次 ~~~~~

|                                                                      |            |    |
|----------------------------------------------------------------------|------------|----|
| 巻頭言 役に立つ情報知識学 .....                                                  | 藤原 譲 ..... | 1  |
| 論文 大規模ヒト 3 次元脳画像データベースの構築<br>志田 和人、川添 良幸、桑川一也、川島 隆太、大槻 昌夫、福田 寛 ..... |            | 2  |
| 論文 材料設計のための結晶データマイニングへのアプローチ<br>ユーリ・コトリャーロフ、岩田 修一 .....              |            | 12 |
| 抄録 .....                                                             |            | 25 |
| 第8回研究報告会抄録 .....                                                     |            | 26 |
| 研究部会報告 .....                                                         |            | 30 |
| XML解説 .....                                                          |            | 36 |
| 書評 .....                                                             |            | 42 |
| 用語解説 .....                                                           |            | 43 |
| お知らせ .....                                                           |            | 45 |
| 投稿の手引き .....                                                         |            | 50 |

~~~~~



情報知識学会

役に立つ情報知識学

藤原 譲

情報知識学会副会長、神奈川大学理学部

以前に本誌(1999, Vol.9, No.1, p.13-23)で情報知識学に関する研究について、自然科学的に整理すると(1)情報知識の特性解析などの実験的研究、(2)それから得られる情報知識の意味表現、記憶構造、知能機構、などの原理、モデル、アルゴリズムなどの理論的研究、および(3)次世代計算機や各種高度実用システムなどへの応用研究・技術開発などの3領域に分類できることを述べた。今回はそのフォローアップの一つとして情報知識学会が実際に役立つものであり、多方面での新しい応用が展開できるのでその一例を述べることにする。

情報化の進展により生活の便利さが増し、また情報の利用範囲の拡大もともに急速でありこれまで計算機や通信にあまり関心のなかった人も情報の機器を使わざるを得なくてきており、高校での授業にも正式に取り入れられることになった。小学校、中学校でもすでに実験段階を済ませ実施の段階であることもよく知られているとおりである。高校での授業とともにそのための教職課程も正式に13年度から始まることになり、遅れ馳せながら体制が整うことになったのは喜ばしいことである。

しかしながら教育を行うことになると対応する専門知識の整理ができていることが必要であることは説明するまでもないことであるが、情報に関しては大学においても何が基本で何が重要であるかが体系的に教えられていないのが現状である。勿論それぞれのところで米国の計算機学会 (Association for Computing Machinery) やそれに準じた日本の情報処理学会の標準カリキュラムをベースにしているが、それには情報学基礎論や知識構造

学といったコースは含まれていないのでどこでも教えていないことになる。誤解をさけるために付言すると、よく知られているようにデータベースや人工知能でデータモデル、知識表現はこれまでに非常に沢山の研究がなされ階層型、ネットワーク型、関係型、述語論理型などなど処理、管理の観点からなされた研究とその成果としてのシステムは数多く報告され市場に提供されている。しかしそれらは情報の特性から要求される充足条件を十分に考慮したものではないので、いずれも重大な問題を回避したままで特定対象に使うか、またはほとんどの場合全く使われないままになっている。

勿論計算機や通信の基礎からプログラミングやシステム構築を含めて計算機や通信システムを使ったり、作ったりすることに関しては基礎から応用までの科目が揃っているのが現在使われている Turing マシンの範囲内のことはやっていけるわけであるが、意味処理に代表される内蔵プログラム主導の方式の枠を超える問題にたいしては展開の見とおしを得ることが困難なままである。情報知識学会は関係機関と共同して、これからの情報化社会を支える情報の知識資源の構築と、それに基づき情報を個人の好みや能力に応じて楽しく学べる役に立つ教材の作成プロジェクトを提案する次第である。なお、他の分野において同じような方法で教材を作成することもある意義であると思われる。

論文

Construction of a large scale 3D image database of human brain

Kazuhito Shida[†], Yoshiyuki Kawazoe[†]
 Kazuya Kumekawa^{††}, Ryuta Kawashima^{††}
 Masao Otsuki^{†††}, Hiroshi Fukuda^{††}

[†]TAO, Aoba Brain Imaging Research Center, IMR Tohoku Univ

^{††}School of Medicine Tohoku Univ

^{†††}IDAC Tohoku Univ

Nowadays, highly advanced medical imaging devices, like MRI(Magnetic Resonance Imaging), are developed and have become rather common apparatus for medical research of the human brain. No matter how much data is obtained, however, it is nothing without appropriate measure for searching and analyzing it. The necessity of sophisticated database management systems for three-dimensional brain images is now clear.

A three-dimensional brain image database which consists of about 1000 brain images taken from normal subjects are under construction. In this paper, what have accomplished on the database up to now will be introduced and several issues on general design of such kind of database are discussed.

1 Introduction

Development of various medical imaging devices enables us to obtain vast amount of three-dimensional data on the final frontier of modern science: the human brain^[1]. The vastly accumulated data is, however, nothing without appropriate measure for searching and analyzing. The necessity of sophisticated database management systems for three-dimensional brain images is now clear.

The brain images, each of which is 4 to 16 megabytes large depending on format and resolution, should be easily selected and transferred via network and should be delivered to the desktop of the researcher dealing with brain image analysis. Once such a system is realized, the system will be utilized for constructing standard brains and clinical processes which are much helped by real-time searching of the brain images.

We have been constructing a three-dimensional brain image database which consists of about

1000 brain images taken from normal subjects. In the following sections, what we have accomplished on the database up to now will be introduced.

2 Requirement for the database

These point should be satisfied by such database systems.

2.1 Fast reaction speed

The large size of three-dimensional brain image is a challenging factor for present fast network systems. This situation is made worse by huge accumulation of data, because the necessity of mass-storage devices such as automatic tape/disk libraries are eminent. The robotics system which exchanges media give rise to further latency of data delivery and its complicated dependency on access pattern. Fast network itself, e.g. ATM or FDDI will

exhibit several problems due to the large size of files to be transferred.

To study this issue, we have employed an ATM network which is spread among several medical institutes and is rather heterogeneous^[2]. Also we have a CD-ROM based mass-storages device. We are trying to check how efficient the data distribution can be on this hardware and to improve it.

2.2 Extended data-usability

In these kind of database systems, the idea of data-usability should be far more extended than they are in the conventional databases. The theoretical background of the relational database management systems(RDBMS) is proven to be very useful to organize various kinds of complicated data structure. It seems, however, that the RDBMS does not offer much help in the case of image database. This is especially true when the organization of the images are flat and relationship between images are mainly defined on-the-fly manner, as in the case of ours.

Flexible image handling and easy access to the image processing tools gives far larger impact on data-usability. Also, since the database will deal with three dimensional images, which is hard to grasp for human user, the idea of data-usability should include accessibility to image visualization tools from the database. Otherwise, the users have to invoke image tools separately from database interface, hence the integrity of the whole system will be far degraded.

2.3 User-friendly interface

The system does not have to have so called "user-friendly" interfaces because all supposed users in this case are professional researchers or medical doctors. Nevertheless, detailed knowledge on the structures of the database should be hidden to users who are busy in their medical works. Nowadays, even the

Structured Query Language(SQL)^[3], which was originally designed for non-professional database users, is too complicated and assuming too much background on the structure of the database.

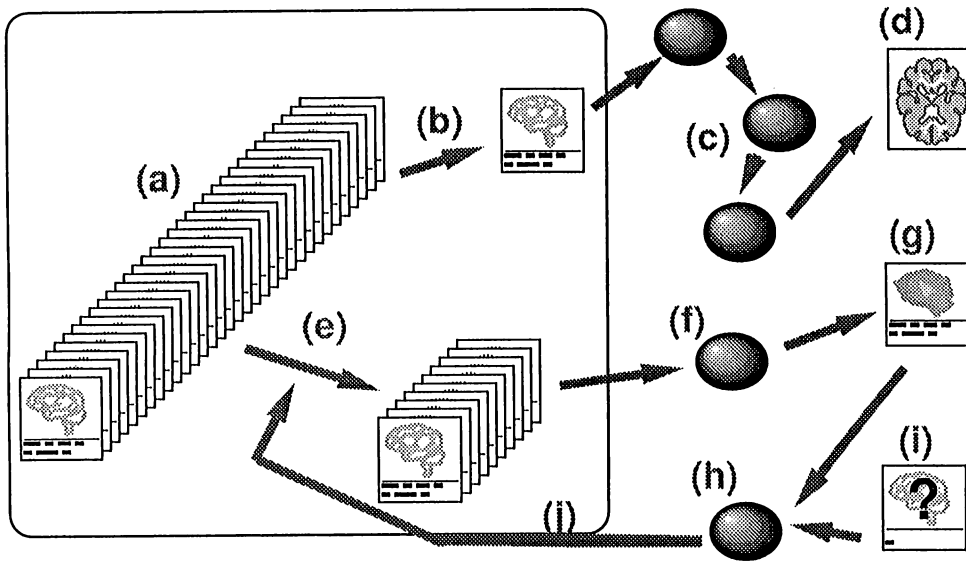
The World Wide Web(WWW)^[4] guarantees an easy and graphical access from database to almost any kinds of client computers. And related technologies like the form interface and accompanied Common Gateway Interface(CGI) will be a nice alternative interface. The form interface is very intuitive and hiding details of the system well.

However, the form interface has a fundamental limit on the complexity of search conditions. Moreover, users of image databases tends to use complex search conditions in combination with various image processing procedures which might be under development and undergo daily changes.

In Figure 1, such a situation is exemplified.

In the figure, the large rectangle at left of the figure represents the database system. Small rectangles are original and processed brain images. And elliptic objects are representing various image processing tools. A series of image processing should be applied to obtain typical representation of a image from the database ((a) in the figure) like a slice-view ((b),(c),(d) in the figure). As shown in the following section, the standard brains are calculated from certain group of images((e),(f), (in the figure)). Such procedure might be repeated with different parameters as a kind of optimization process, ((h),(i),(j) in the figure) as discussed in the following section.

In these examples, the image data handling, the image processing and the handling of their result are so intertwined and kind of algorithm should be involved to perform the whole process. Character based interface like some kind of shell or command-interpreter will be more appropriate in such cases. Also, a script language function may be provided by such shell, which might be nice medium on which many other tools communicate each other. This will give rise to the above-mentioned



integrity of the image processing system.

3 The standard brains

We are especially interested in processing these brain images from normal subjects because the brain images without any lesion or defect must keep their natural forms. From natural forms, we can derive an average shape of the brain. Such an average is referred to as a standard brain. A standard brain is very useful for computational analysis of brain images because it gives us an anatomical mapping on the brain.

Also, such an average shape of brain might be useful for clinical application. If there is some statistically significant difference exists between a patient's brain and the standard with all condition (age, sex, etc.) the same as the patient, it might be understood as a sign of diseases.

In other words, by gathering normal brains, standard brains are able to be defined for each sub-set of the brain images. Image sub-sets consist of brain images from subjects who have any special attribute such as male, female, special metabolic problems, in common. Then we can discuss several issues by means of the standard brains.

Firstly, when a new brain with known attributes is given, the standard brain of the sub-set with the same attributes can be used as template of pattern recognition to identify each part of the brain. Such process is called registration in the field of medicine. The parts of the brain to be identified are, namely, the frontal lobe, temporal lobe, occipital lobe, and parietal lobe. These are corresponding to front, side, top, and dorsal parts of the brain, respectively. And more detailed registration can be done in similar manner. Using finer subset, that is, specifying more conditions for sub-sets which standard brain based on, is effective to improve precision of the registration. Today, automated registration is highly advanced^[5], but it still is imperfect and often needs human supervision or final check on the result. Hence, the result, in a form of color-coded image, should be related to and organized with the original image in the database.

Secondly, when a newly given brain's attributes are unknown, the difference from the standard brain will be helpful to decide whether it should be included to the sub-set. This will be helpful for clinical diagnoses by taking the standard brain of the healthy subset. Specifying more conditions give rise to reliability of diagnosis. For example, the reduction

of brain volume, which is not rare in elderly people, may be diagnosed abnormal taking a standard brain of whole population as the benchmark. This can be avoided by using a standard brain defined for people elder but proven as healthy. In such a task, brain images which satisfy certain conditions should be handled as a group, so that researcher's workload is minimized.

Thirdly, when standard brains are defined for more than one sub-sets on which we expect some meaningful difference, comparing the standard brains of those sub-sets are best way to evaluate the difference of the sub-sets. Also, for sub-sets parametrically classified, such a procedure is useful for parametric understanding of the difference. This is especially interesting when the parameter is the age, because this is verily clarifying the mode of aging of the brain, which is very important theme of medicine in the near future. However, to discuss effects of various factors on retarding or aggravating the aging, the standard brain should be made for various sub-sets on demand. In this case, if calculating a standard brain for particular subset can be initiated by single command, it is the ideal way of introducing abstraction to database. Again organization of images and image processing tools is important.

4 Source of the brain images

As explained above, the normality of the subjects is especially important. All brain images should be free from defects. All brain images incorporated into the database should be taken from subjects who are free from diseases which can affect the shape of the brain. At the same time, the number of the subject is also important for reducing the statistical error of the standard brain. As subjects, we have been inviting volunteers.

Having noticed that considerable number of the volunteers who claim to be healthy by themselves have severe diseases, we are carrying out the following screening procedure

for every one of the volunteers.

After the agreement to conditions of our research project and confirmation of ones "informed consent", every one of volunteers is examined by a doctor of our team, and a set of MRI images of ones brain is taken. The data obtained from examination process is recorded in a form of questionnaire. This data will be introduced later as attribute data of the brain images. The doctors of our team check the state of the subjects by means of the MRI images and classify them into following five classes.

- A. No lesion is recognized in their brain images.
- B. Limited number of lesion are recognized.
- C. Worse than B. However, practically they are health.
- D. Severe problem is found in their brain.
These subjects are introduced to clinical doctors. Since all our procedure is just for quick screening of unhealthy brains, further formal examination is recommended for subjects classified to this class. They are not included in our data base.
- O. The brains in this group are actually belong to one of A, B, and C, class. However, some problem is found in the MRI images, somewhere but in the brain. Namely, they have problems in their ears or nose. These subjects are introduced to the otorhinolaryngologists.
- X. There is a problem on quality of the MRI images. Possibly a mistake in the scanning. Occasionally, foreign objects, especially metal crowns of dentist on their tooth cause this kind of problem by its large magneto-susceptance. If such foreign objects can not be removed easily, the subject is simply released.

The brain images classified to A, B, and C, class will be incorporated to the data base.

5 Organization of the database

The brain images are stored into the database with their attribute data. The images are addressed and searched by using these data as keys.

Searching for images taking an image as a key, that is, automatic filtering of images based on similarity between the images and given key, is recently getting large attention. However, in our project, such kind of image searching is kept as a theme for future extension. Because image key is useful only if there are several subsets of images which are unlike to each other. Since all images in our database are normal brains hence the difference between them should be small.

The data model so far we have been using is a relational one. Its main key is the identifiers or the handles of the images related to various attribute data stored in the single table. There are other tables in the system (e.g. for user administration), in this paper they are not introduced for simplicity.

The following are the name of each field of data and its explanation. Some of them are directly corresponding to single SQL field. Others are corresponding to multiple SQL fields. These correspondence rules are also introduced. The datatype (float, character etc.) is decided to maximize the data usability of the completed system.

Some of the fields are directly related to particular diseases. We have selected these diseases according to our special interest on them and a natural assumption that only limited diseases have significant effects on shape of the brains.

5.1 Generic data

Subject ID: Consists of unique number and name of the institution the subject is accepted. The latter part is necessary because we are planning to gather MRI images from more than one institutes.

Age: An integer.

Sex: A string field contains either of "Male" or "Female".

Height: A floating number in centimeter.

Weight: A floating number in kilogram.

Blood pressure: Two floating numbers called BP_L and BP_H. They are translated from widely used notation like "80/170". The unit used for BP_L and BP_H is "mmHg".

Handedness: A floating number between -1.0 and 1.0. This value is the likelihood of the subject's handedness: -1.0 means completely left-handed. 1.0 means completely right-handed. This value is evaluated by a standard test^[6].

MRI diagnose: The diagnosis made by doctor by means of the checking in advance which is described in chapter 4.

5.2 Life style

Smoking: Whether the subject ever have a smoking habit. Either Y or N.

Drinking: Whether the subject ever have a drinking habit. Either Y or N.

Smoking index: A floating number which describes how much is the subject currently smoking.

Drinking index: A floating number which describes how much is the subject currently drinking.

Profession: Current profession of the subject. A string field.

Exercise: What kind of exercise the subject is currently doing. A string field.

Hobby: What kind of hobby the subject is currently doing. A string field.

5.3 Psychological test

PSYCOTEST: Either Y or N, meaning whether the subject has cooperated in the psychological test. This field is necessary because the subjects are allowed to deny this part of examination if they do not want.

PSYCOVAL: Twenty-five floating values each of which are representing specific trend of the subject's psychology. Originally extracted from a standard test^[7].

5.4 Current Diseases

For each of selected diseases, there are three fields describing whether the subject has the disease (Either Y or N), how long they have been with that (A floating number in years.), and what kind of medication or treatment they have undergone (a string field). Selected diseases are:

- Hypertension
- Hyperlipoidemia
- Hypercholesterolemia,
- Diabetes

5.5 Past Diseases

Any diseases the subject have experienced, then recovered. A string field.

5.6 Past Diseases on Family

For each of selected diseases, there are five fields (Either Y or N) each of which represents whether the subject's parents, sibling, grandparents, children, and grandchildren have or had the diseases. Selected diseases are:

- Hypertension
- Hyperlipoidemia

- Hypercholesterolemia
- Diabetes
- Cerebral infarction
- Cerebral hemorrhage
- Subarachnoidal hemorrhage
- Psychiatric disorders
- Arrhythmia

5.7 Subjective symptoms

Headarch: Whether the subject is complaining headarch. (Either Y or N)

Vertigo/Dizziness: Whether the subject is complaining vertigo or dizziness. (Either Y or N)

Insomnia: Whether the subject is complaining problems on sleeping. (Either Y or N)

Ear ringing: Whether the subject is complaining a ringing in ears. (Either Y or N)

Forgetfulness: Whether the subject is complaining forgetfulness. (Either Y or N)

Auxiliary: Any complaint the subject makes. (A string field.)

5.8 Current usage of drugs

For each of selected kinds of drugs, there are three fields describing whether the subject has been taking the kind of drug (Either Y or N), how frequently they are taking that (A floating number in times per week.), how long they are taking the kind of drug on continuous basis (A floating number in years.). Selected kind of drugs are:

- Sleeping drugs
- Drugs for headarch / painkillers

Table 1. Subject statistics

	Male	Female
10-19	24	5
20-29	74	38
30-39	35	25
40-49	58	57
50-59	85	103
60-69	43	33
70-100	11	8

- Traditional Chinese medicine
- tranquilizers

6 Current status of the database

Up to now, we have successfully gathered over 600 acceptably healthy brains among all age group of Japanese, and construction of the average brain is in progress.

The age/sex distribution shown in Table 1 is a little bit unbalanced. We are now trying hard to fix the balance so that the statistical error of standard brain is minimized.

We have simulated the situation which the full-scale database will face to, and checked the reaction time of search and image retrieval in the case of ATM-connection. The data file is $256 \times 256 \times 124$ voxels large and each voxel has 16-bits (16252968 bytes). A typical search condition for single image retrieval was entered to the web-based interface from a ATM-connected client machine to simulate a user's action. The wallclock-time consumed for completing the downloading is measured and shown in the left hand side of Table 2. We also measured the reaction time for single media exchange by our robotics device and included it in the left hand side of Table 2, which gives the worst case reaction time for single image retrieval on the system. We can conclude the performance is enough for real-time image handling on ATM-connected clients especially with compression of the data files. However, access

Table 2. Time for completing operations

	Downloading	Invoking image software
Searching	4.0 Sec.	4.0 Sec.
Media Switching	4.0 Sec.	4.0 Sec.
Transferring	33 Sec. (12 Sec.*)	
Starting software		4.0 Sec.
Sum.	37-41 Sec. (16-20 Sec.*)	8-12 Sec. 8-12 Sec.

*: data compression by "gzip -9".

to multiple images will cause another negative impact on the performance via switching of medium. We have devised a novel way to alleviate these problem [8] and introduced a simulation result of performance improvement.

As the first step of database construction, we have extracted about 100 brain images out of 600, and constructed a prototype database environment with minimum functions on a UNIX database server (DEC Station 4100 with 4CPUs) and a IBM-compatible type personal computer.

Following the above mentioned requirements for the interfaces, we have been developing two different prototypes each addresses different aspects of the requirement for interface. The web-based interface (Figure 2) provides the basic level accessibility and usability to the database[9]. It is connected to a prototype database as the backend. The backend is based on the Informix IUS 9.12 RDBMS working on a IBM-compatible type personal computer as well as the web-browser working as the interface. The search condition can be given to any combination of the database fields in the table-like interface. The web interface can deal with logical combinations of the search criterion by and/or selector in the rightmost column of the table. Once some of the images are selected, the user can specify local directory into which the selected im-

ages will be downloaded in a download dialog. Later, user may invoke any processing tools on the client machine.

The other interface we are developing is a character-based one. Actually, it is a programming environment on which users can execute a kind of script languages, currently an extension of the Perl5^[10]. Because this is a kind of rapid-prototype, the key issue is the simplicity. Hence, the backend database is implemented on Perl5 as well as the interface. The database tables are stored into hash valuables of the Perl5. And, some of the functions like user administration and security issues are largely exploiting the UNIX environment's native functions. The system is implemented on the DEC Station server, which turns out to be fast enough to cover the efficiency problem of the interpreter language.

The main characteristics of the environment are,

- In general, object-oriented design is employed. By this, a modular design is enabled, and future extension will be easier.
- A search function by which users can filter the images by any condition on attributes of the images. This also provides full-text search on contents of string type attribute fields.
- Result of the search function can be substituted in Perl valuables. This enables utilizing the result for adjusting the condition of the next search. Moreover, it is very convenient to address a group of images by the name of single Perl array.
- A help function using full-text search on a repository. By this function, character-based interface can be even more user-friendly than web-based one.
- An extended history function, which can keep track of any file generated in the

process of image handling and analyzing, is under development. The efficiency of development of image processing on trial and error basis will be improved because the trace of such a task is automatically recorded.

- Since the environment is based on the Perl5, various tasks on the database can be executed by simple scripts which are directly executed by the Perl5.
- Image processing tools can be invoked directly from the environment. This gives rise to further integrity of the database image processing systems.

On such an interface, the user can implement rather complicated statistical operations. For example, an algorithm of cluster analysis of brain images may be implemented as a script in combination with the image processing tools, which can evaluate the similarity of brain images given by the database.

The reaction time of this interface is also measured and presented in the right hand side of the Table 2. In this case, retrieved image is not stored into the harddisk of the client machine, but is directly loaded by the image processing program to be presented in the display of the client machine. It is impressive that reaction time is largely improved by this way of handling images.

7 Conclusion

Toward the future research of human brains, construction of a large scale brain image database is on its way. To clarify problems of the project, a prototype database was constructed and its basic performance was measured. Also, several issues on preferable interface of such kind of database is discussed.

8 References

- 1 "Brain Mapping : The Methods" A. W. Toga (Editor), J. C. Mazziotta (Editor)

検索条件フレーム - Microsoft Internet Explorer

アドレス http://demo/cgi-bin/webdriver.exe?Mlval=CordFrame&userid=system&allg=10

検索項目および条件を指定してください。

検索項目をチェックし、項目の条件を選択または入力してください。
検索項目がチェックされていないものは検索条件にはなりません。
また、一項目もチェックされていない場合は全件検索になります。

項目名	条件	演算子
<input type="checkbox"/> 映画(演劇映画)	<input type="text"/>	C And C Or
<input type="checkbox"/> 性別	<input type="checkbox"/> 男 <input type="checkbox"/> 女	C And C Or
<input type="checkbox"/> 年齢	<input type="checkbox"/> オ <input type="checkbox"/> 才	C And C Or
<input type="checkbox"/> 身長	<input type="text"/> cm <input type="text"/> cm	C And C Or
<input type="checkbox"/> 体重	<input type="text"/> kg <input type="text"/> kg	C And C Or
生活習慣		
<input type="checkbox"/> 喫煙歴		C And C Or
<input type="checkbox"/> 飲酒歴		C And C Or
現病歴		
<input type="checkbox"/> 高血圧	期間 <input type="text"/> 年 ~ <input type="text"/> 年	C And C Or
<input type="checkbox"/> 糖尿病	期間 <input type="text"/> 年 ~ <input type="text"/> 年	C And C Or

検索ボタン

Academic Press (1996)

- 2 "A Guide to the SQL standard" C.J.Date.
Addison-Wesley (1987)

- 3 K. Kazuya, K. Shida, R. Kawashima, M.
Otsuki, Y. Kawazoe, H. Fukuda. IEICI
abstracts: (1998)

- 4 <http://www.w3.org/>

- 5 D.L. Collins, C.J. Holmes, T.M. Peters and
A.C. Evans. Human Brain Mapping pp190-
208 Vol3, (1995)

- 6 R.C. Oldfield. Neuropsychologia 9,1, pp97-
113 (1971)

- 7 P.T. Costa, R.R. McCrae Journal of Per-
sonality and Social Psychology 55,2, pp258-
265 (1988)

- 8 K. Shida, K. Kumekawa, R. Kawashima,
M. Otsuki, Y. Kawazoe, H. Fukuda IE-
ICI abstracts. (D-4-7) (1999)

- 9 K. Shida, K. Kumekawa, R. Kawashima,
M. Otsuki, Y. Kawazoe, H. Fukuda IE-
ICI abstracts. (D-4-3) (1999)

- 10 "Programming Perl" L. Wall, T. Chris-
tiansen, R.L. Schwartz, S. Potter O'Reilly
and Associates Inc. (1996)

(1999年12月5日受付)

(2000年7月21日採録)

著者紹介

志田 和人

北陸先端科学技術大学院大学 情報科学
研究科 博士後期課程終了。
現在、通信放送機構青葉リサーチセンタ
研究員。
脳画像 DB 開発に従事。
Email: shida@art.imr.tohoku.ac.jp

川添 良幸 (正会員)

東北大学理学研究科 核物理学専攻 博士
課程修了。
理学博士。
現在、東北大学金属材料研究所教授。
電子情報通信学会、情報処理学会、日本
金属学会、MRS など各会員。

桑川一也

東北大学大学院 理学研究科博士課程修了。
現在、電気通信大学 総合情報処理センター

助手.
通信ネットワークの研究に従事.
日本物理学会、人工知能学会、日本核医学会各会員.

川島 隆太

東北大学 大学院医学研究科 博士課程修了.
医学博士.
現在、東北大学 加齢医学研究所 専任講師.
医学博士.
主としてPETを用いた脳機能マッピングに従事.
米国神経科学会、日本神経科学会、日本生理学会など各会員.

大槻 昌夫

東北大学医学部医学科卒業.
医学博士.
現在、東北大学医学部 附属病院医療情報部 教授.
慢性肝疾患の病理形態的研究、ウイルス性肝炎の治療.
教育病院における医療情報に関する研究に従事.

福田 寛

東北大学 大学院医学研究科 博士課程修了.
医学博士.
現在、東北大加齢医学研究所 教授.
非襲性生体画像装置による脳機能の解明、癌診断法の開発研究に従事.

論文

Layers Sequences in the Materials with Perovskite-Related Structures

Yuri Kotliarov[†], Shuichi Iwata^{††}[†]Research into Artifacts, Center for Engineering, the University of Tokyo^{††}Institute of Inorganic Chemistry, Siberian Branch, Russian Academy of Sciences

A new approach to get a qualified view on large amounts of materials data is proposed. Global understanding on the relations among chemical composition, crystal structure and properties of materials on the basis of suitable classification is a starting point of material design. The present work is focused on studying the crystal structures and structural primitives as keys for materials design based on crystallographic information, and a sequence of stacking 2-dimensional layers, implying hidden regularities in the high temperature superconducting (HTSC) materials, is selected as an additional features to generic crystallographic information.

Prerequisites for using this approach, its limitations and possibilities for database developments are discussed based on the analyses of data extracted from Inorganic Crystal Structure Database (ICSD). In addition, possible applications of this method for classification and design of these interesting materials are proposed.

1 Introduction

The more we get the insight on the semantics of material structures, the more we can extract the rich properties from materials. And it is the first milestone to explore the new domains of materials in the process of materials design. In this paper, an approach to describe material structures beyond traditional crystallographic and metallographic approaches is proposed to have a new articulation on material structures by taking advantage of fascinating high Tc superconducting materials.

The presence of CuO₂ conduction sheets is crucial for the presence of superconductivity in copper oxide compounds. Their overall structures are usually described as stacks of 2-dimensional sheets and structural blocks along the direction perpendicular to the sheets. As all these compounds comprise similar structural units, describing their structures in a coherent manner becomes useful for clarify-

ing their common structural features and relationships among various crystal structures. Such a structural representation also serves as a tool to devise strategies to search for new compounds that may exhibit interesting electronic and magnetic properties.

The four number classification scheme of superconducting copper oxides with very close relation to their layered structure has been proposed by Shaked^[1]. However, this approach, although gives us very convenient representation of the material, can not adequately describe its crystal structure.

Tokura with his colleagues^[2] has proposed another classification method for layered copper oxide compounds. Every HTSC compound possesses alternately one-to-one stack of CuO₂-sheets and "block layers", or intervening atomic layers which are composed of metal ions (mostly alkaline earth and/or rare-earth ions) and oxygen or halogens. All CuO₂-layered compounds are classified into the combination of "block layers". Moreover, the

"block layers" sequence governs fundamental physical chemical properties of the compounds, such as crystal structure (including space group), Cu-O bond lengths and prescriptions for doping by p and n type carriers.

A convenient method of expressing the (ideal) structure of superconducting copper oxides by regarding them as combinations of perovskite and rocksalt-type sheets has been proposed by Santoro et al. in 1988^[3]. Although perovskite-like and related commensurate structures (like rocksalt or fluorite) were considered as layered and described as stacking layers even before the HTSC discovery^[4], their work^[3] was the first systematic analysis of layered structures of copper-oxide superconductors from the viewpoint of layer patterns and stoichiometry. Later the development of this approach was continued by the same authors^{[5][6]} and also in other publications^{[7][8][9]}.

The present work at the first time considers this approach as an advanced "data mining" procedure to search for compounds with similar but different crystal structure. When we are going to find compounds with closely related structures, such as homologous series members, solid solutions, derivative structures, etc., standard principles based on space group theory cannot be very helpful. Disadvantage of such methods is that even a minor change in the position of the atom in a crystal structure can often reduce its symmetry, leading to a new mathematical description and a "new" crystal structure. Instead of principles of symmetry, rearrangement of atoms in crystal structure and structural primitives should be taken into account. Following this idea, the analysis of layered structures of copper-oxide compounds, with huge amount of representatives, can be a good example for such a procedure. In addition a special interest is paid to their superconducting properties, and this interest does not decrease during last decade. This paper discusses the problem

of automation of sequence of layers building procedure for "data mining", classification and comparison of different structures, and for other purposes.

2 Methods

In the present work widely famous Inorganic Crystal Structure Database (ICSD) has been used as a data source^{[10][11]}. CD-ROM version of the database contains about 37.5 thousands of entries of different inorganic structures.

Software programs have been written on C language (ANSI standard) with the help of SgInfo comprehensive collection of routines for handling of crystallographic space groups^[12].

3 Basic definitions of the layers approach

The layered structure can be described in different ways and at different levels of details. Perovskite-like structures, despite of the great variety, allow to describe them with the help of small set of atomic networks perpendicular to (pseudo)-four fold axis⁷. The layer composition may be expressed by the chemical formula corresponding to the stoichiometry of the layer, for example A, AX, B₂, BX₂, etc. For the sake of simplicity, generally atomic positions are redefined as A - large size cation position, B - small or middle size cation position, X - anion position. Subscript indices are shift of cell relatively to coordinate origin with translation vector: o=(0,0); c=(1/2,1/2); x=(1/2,0); y=(0,1/2). Square brackets, using in some formulae, indicate the layers that are enclosed in one unit cell of the structure. This notation, which is more close to Cava⁵, has been chosen because of its simplicity for automation. In other works notations are a little bits different. Furthermore, each pattern can be represented by a digital code. Although this code does not describe atomic arrangement, it sig-

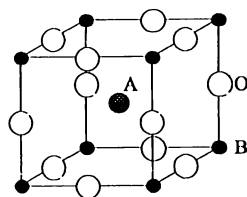


Figure 1a Schematic representation of crystal structures of perovskite

nificantly simplifies the computer analysis of the patterns.

For example, if we apply this symbolism to the structure of perovskite CaTiO_3 (Figure 1a), we obtain the sequence $[(\text{BX}_2)_o(\text{AX})_c]$, where $\text{A}=\text{Ca}$, $\text{B}=\text{Ti}$ and $\text{X}=\text{O}$; or simply -1o5c (see Figure 2). The rocksalt structure, the second main building block in cuprates, is represented as sequence $[(\text{AX})_o(\text{AX})_c]$ or 5o5c. For Y123 structure (Figure 1b) the sequence will be $[(\text{BX}_2)_o(\text{AX})_c(\text{BX}_2)_o(\text{A})_c(\text{BX}_2)_o(\text{AX})_c(\text{BX}_2)_o]$ or 1o5c1o8c1o5c1o. As one can see, the perovskite block in Y123 phase can be identified easily with the help of the sequence of layers. If necessary (for example, for the analysis of formal electronic state of elements in layers), letters A, B and X can be replaced back by real symbols of the elements, as $[(\text{TiO}_2)_o(\text{CaO})_c]$ for the perovskite structure.

4 Possibilities for automation of layers sequence building process and pattern matching

The procedure of automatic analysis of layers and building them into the sequences requires the primary definition of limiting amount of basic elements - layers patterns, which would cover majority of possible copper-oxide structures. Here the common building principles of such structures are considered in context of the possibility of revealing basic structural primitives and limiting their amounts.

The basic cell for such types of phases is a planar network of CuO_2 composition with

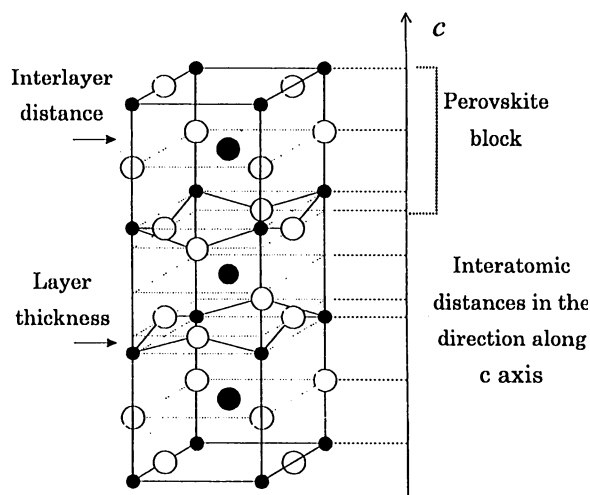


Figure 1b Superconducting $\text{YBa}_2\text{C}_3\text{O}_{7-x}$ (Y123 phase) with perovskite blocks.

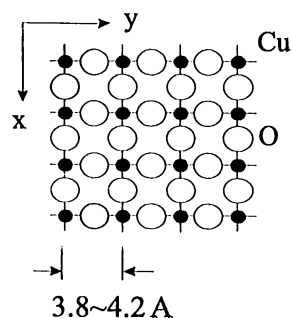


Figure 1c Basic fragment of layered perovskite-like structures Cu-O network

lattice parameters $a \sim b = 3.8 \sim 4.2 \text{ \AA}$ [13]. Its schematic representation is shown in Figure 1c. At such short values of lattice parameters the number of possible cation-anion cells, proportionate CuO_2 , is limited - only nine topologically different patterns. The networks BX and $(\text{BX})'$ are topologically identical, but differ by rotation by 90° in xy plane.

Figure 2 illustrates schematic representation of possible atomic networks for perovskite related compounds.

Short interatomic distances ($2.7\text{--}3.0 \text{ \AA}$), approximately equal to $a/\sqrt{2} \approx b/\sqrt{2}$, on the other hand, seem to play a role of lower limit of cation-cation or anion-anion bonds (if compounds with chemical bonds of the corresponding types are not taken into ac-

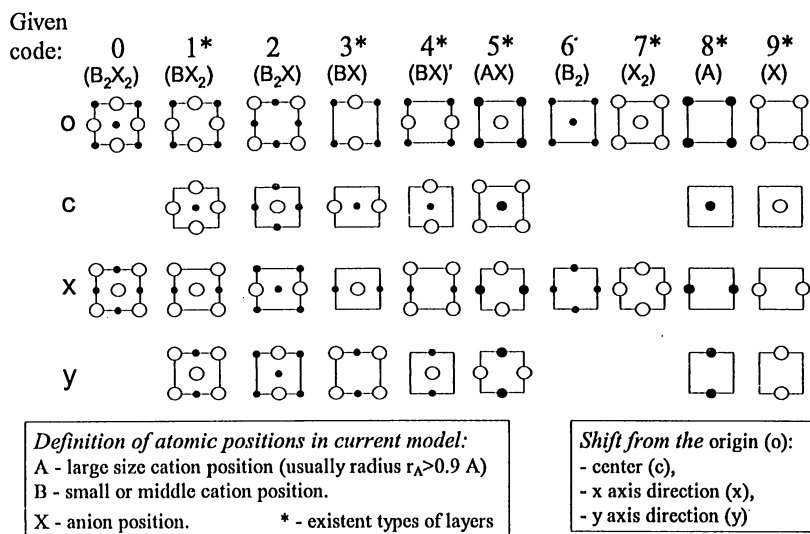


Figure 2 Possible atomic networks for perovskite related compounds

A - large size cation position; B - small or middle cation position; X - anion position. * - existent types of layers. Shift from the origin (o): c - center, x - x axis direction, y - y axis direction.

count). On the other hand, it is approximately equal to the upper boundary of the shortest lengths of cation-anion (metal-oxygen) distances. Therefore, this number specifies both the maximum distance between layers (in copper-oxide HTSC the maximum inter-layer distance does not exceed 2.4 \AA ⁷) and certain limitations on their combination, preventing from neighboring atomic networks, identical by type and shift (with the same subscript symbol). Similar steric reasons prevent large cations from allocation in B₂X₂, BX₂, B₂X, BX, B₂ and X₂ networks, that uniquely determines X as oxygen or fluorine and ionic radius $B^{n+} r_B < 0.9 \text{ \AA}$. Some variants of allocation of cations with different charge in the networks cannot be realized also because of limitations of electrostatic character (improper charge balance), resulted from Pauling's electrovalence rule¹⁴. Summarizing the above statements, practically all the possible combinations of considered networks could be defined in advance. However, as will be shown later, most of them

do not exist or very rare in practice (Table 1). It can be claimed that the layer formalism considered with the data from Table 1, allows complete automating of the process of modeling perovskite-related structures.

However, the method of identifying individual layer in perovskite-like superconductors is not obvious, in contrast to, for example, the clearly expressed layered crystals of the type of As₂S₃, GeSe, and others. In the latter compounds the bonding of atoms within the layers has a covalent nature, and the interaction between the layers is determined by van der Waals forces.

Developing the layer pattern matching we need to solve the following problems.

- Orthorhombic distortions. Different orientations of unit cell.
- Superstructure and modulation.
- Complex anions, e.g. carbonate CO₃²⁻.
- Deviations of atomic positions from the ideal model.
- Non-plain layers.

	B_2X_2	BX_2	B_2X	BX	BX'	AX	B_2	X_2	A	X
$(B_2X_2)_o$	X	XY	O	XY	XY	-	X	O	XY	OC
$(BX_2)_o$	X	XY	OC	XY	XY	C	X	O	CXY	OC
$(B_2X)_o$	O	OC	XY	OC	OC	O	O	X	OC	CXY
$(BX)_o$	X	XY	OC	CXY	XY	CY	X	OX	CXY	OCX
$(BX')_o$	X	XY	OC	XY	CXY	CX	X	OX	CXY	OCY
$(AX)_o$	-	C	O	CX	CY	CXY	X	X	CXY	OXY
$(B_2)_o$	X	XY	OC	XY	XY	XY	X	OX	XY	OCXY
$(X_2)_o$	O	OC	XY	OC	OC	XY	OX	X	OCXY	XY
$(A)_o$	X	CXY	OC	CXY	CXY	CXY	X	OX	CXY	OCXY
$(X)_o$	O	OC	CXY	OCX	OCY	OXY	OX	X	OCXY	CXY

Table 1 Possible variants of layers combination, satisfying the crystal chemical rules.

Each cell represents possible variants of shift (see Figure 2) of the column layer relatively to row layer, that is in the origin. Revealed variants are shown in bold. O - origin; C - center; X - X direction; Y - Y direction.

In the superconducting structures, copper-oxide layers tend to be either planar or relatively thin. Other layers such as LaO or BaO are ordinarily much thicker, generally to such an extent that it seems more reasonable to consider the cations and oxygen atoms as forming separate layers. This may cause a situation that we cannot separate each layer from others without a certain ambiguity even if we have a layered structure due to the possibility of intersection of layers. To show the real layered character of cuprates structures, interatomic distances in them along characteristic direction (axis c) have been calculated, that will be described later.

5 Developed programs

As a result of the present work, a set of computer programs for data mining, building crystal lattice and layer sequencing have been developed:

- I. Crystal Unit Cell Builder - program to generate all atoms within unit cell from the basic atomic position table. C routine library SgInfo^[12] has been used for handling symmetry operation and tran-

sition matrices. The program also checks, whether new position located too close to another position, that may be results of atomic deviations and partial occupancy.

- II. Layer Sequencer - program for creating the sequence of layers in crystal structure and for 2-dimensional representation of layers in ASCII graphics. This program is under development and the present version treats only structures with parameters $a \sim b \sim 3.8$ Å. It isolates layers along axis with maximum length and matches them with the patterns given in advance. If the separation of layers or pattern recognition is ambiguous, the program puts question marks in the sequence and allows user to make decision.
- III. ICSD to Table Converter - utility to convert output data from ICSD database to a tab delimited text file, which is used to upload the data to database table.
- IV. Reference Extraction - utility to extract reference information from ICSD database and convert it to *MedLine format which can be used by any reference*

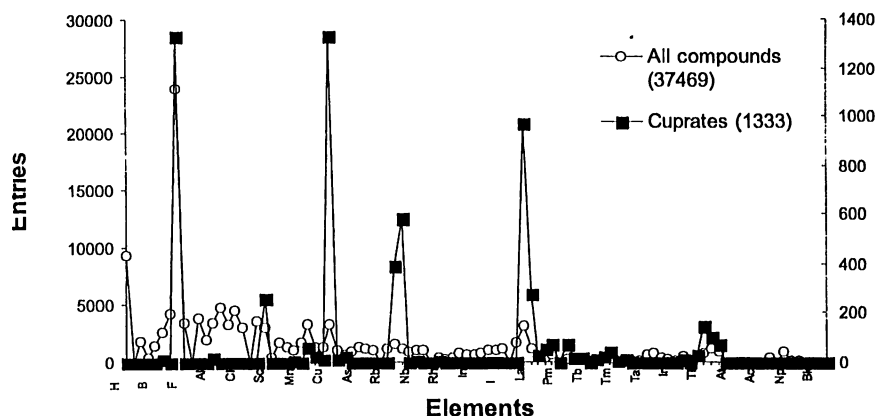


Figure 3 Distribution of compounds extracted from ICSD by element composition

manager program.

6 Preliminary analysis of data extracted from ICSD

A search query is started by identifying chemical components on the basis of knowledge about similarities in superconducting and related cuprates:

(Cu and O) and (Ca or Sr or Ba) and (Y or La or LAN or Hg or Tl or Pb or Bi)

As a result, with the help of build-in RETREIVE v.2.0 software, 1333 cuprates have been extracted from ICSD in this work.

To have a clear view and to show a big picture of the using array of data, distribution analyses of elemental compositions and structural properties have been done as shown in Figures 3, 4 and 5. As it can be seen in Figure 3, close-packing principles in perovskite-like compounds result in very limited set of possible atoms in these structures. Last two figures clearly show that there are not so much varieties in the distribution of these compounds by structural properties: orthorhombic and tetragonal structures are dominant, and only four space groups prevail. This is due to the very strict limitations which CuO_2 network imposes to the crystal structure of these compounds.

As for the space group appearance, it is im-

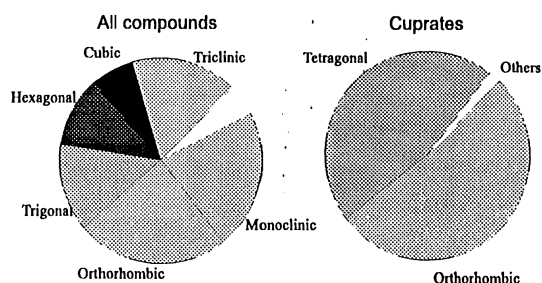


Figure 4 Crystal symmetry population pattern for inorganic materials extracted from ICSD.

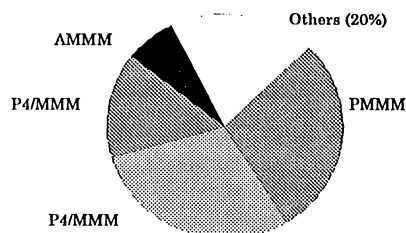


Figure 5 Distribution of cuprates by space groups in copper oxides (totally 59 groups)

portant to bear in mind that the symmetry of the system depends on the dimension of the subgroup of translations, which may be equal to the dimension of space, but can also be less. In general case when there is a lowering of the space dimension and of the dimension of the translation subgroup, the symmetry of the corresponding object will be described by one of the groups, which have been summarized by Zallen^[16]. According to him, layers are systems with two-dimensional translation symmetry and three-dimensional point symmetry, and their space symmetry can be described by one of the 80 three-dimensional doubly periodic (layer) groups. In our case in total only 59 space groups have been found, and 80 % cases are described by 4 main groups: Pmmm, P4/mmm, I4/mmm and Ammm. That can be explained by additional limitations of perovskite structure. Looking the big picture, it should be noted that the structures, for which sequences of layers have been resolved, are described by only 10 space groups, while above mentioned 4 space groups appear in 98 % cases.

As it has been emphasized before, the above structural layers are not ideally plain, but they may have some thickness. In the considered case the interatomic distance means a distance between two neighboring points, produces by projection of atoms to *c* axis (*z* coordinates) as shown in Figure 1b. The layer thickness in this context is a distance between the minimum and the maximum *z* coordinates of atoms of the same layer. The interlayer distance in its turn is defined as a distance between two neighboring layers, exactly between the maximum *z* coordinate of lower layer and the minimum *z* coordinate of upper layer. Results of calculation are displayed in Figure 6.

On this histogram each column relates to the number of those layers (in this case ideal plain layers are considered) which fall into a specific range of interatomic distances. It is clear to see two separated areas on this graph responsible for layers thickness and real in-

terlayer distances. Thus, it can be concluded that in the majority of cases there is no ambiguity to separate atoms into two-dimensional layers by using this characteristic interatomic distances.

7 Results of the layers sequence building and their analysis

The first two programs described above have been used to calculate the sequences of layers for 1333 copper oxide compounds extracted from ICSD. About 19 % (261) of the compounds have been rejected because of large lattice parameters (the cell reduction procedure is not implemented yet), or imperfect initial data (positions of some atoms have not been defined), or high degree of disordering. The sequences of main representatives of perovskite-related materials are shown in Table 2.

Layers sequences are given in two-character system, where the first digit represents a pattern defined in advance, and the second character describes a shift of the pattern from the origin position. More details have been given in Figure 2.

Figure 7 shows the revealed layer patterns and the number of extracted compounds, where these patterns have been found. Pattern 1-(BX₂) (CuO₂) is the basic building network, determining the most common structural features of superconducting and related copper oxides. Pattern 5-(AX) (metal-oxygen) layers take place in building perovskite blocks and play a very important role in the local charge distribution as shown in the next chapter. Layers 8-(A) (more often Y or Ca atoms) is a separator of perovskite blocks. Copper-oxygen chains, also playing an important role in the superconductivity in 123 phases, are represented by patterns 3 or 4. Patterns 3 and 4 (BO) differ from the pattern 1 (BO₂) only by oxygen vacancies in one of the edge and they together represent the Cu-O layers. As one can see in Figure 7, at

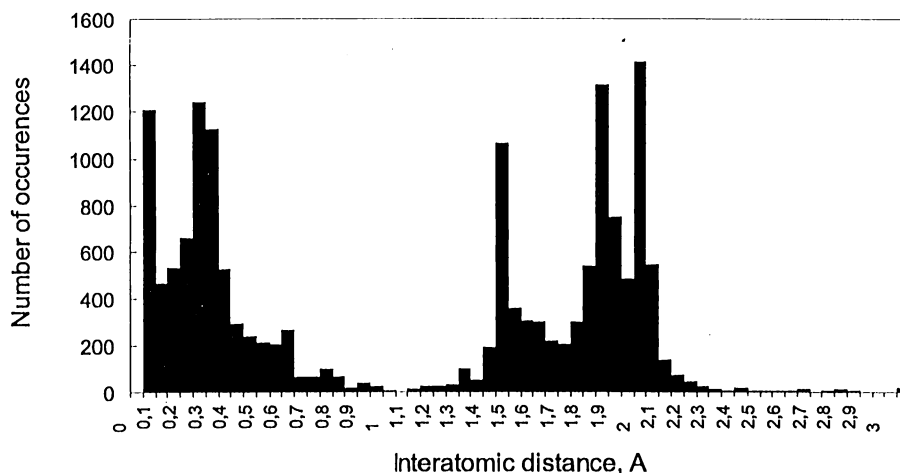


Figure 6 Distribution of interatomic distances along [001] direction (c axis) in cuprates.

N	Compound	Layers sequence	T _c , K
1	(La,Ba) ₂ CuO ₄	1o 5c 5o 1c 5o 5c	38
2	(Nd,Ce) ₂ CuO ₄	1o 8c 7c 8o 1c 8o 7c 8c	24
3	(Nd,Ce,Sr) ₂ CuO ₄	1o 8c 7c 8o 1c 5o 5c	35
4	YBa ₂ Cu ₃ O ₇	1o 8c 1o 5c 4o 5c	92
5	YBa ₂ Cu ₃ O ₆	1o 8c 1o 5c 8o 5c	-
6	YBa ₂ Cu ₄ O ₈	1o 5c 4o 4x 5y 1x 8y 1x 5y 4x 4o 5c 1o 8c	80
7	Y ₂ Ba ₄ Cu ₇ O ₁₄	1o 8c 1o 5c 4o 4x 5y 1x 8y 1x 5y 8x 5y 1x 8y 1x 5y 4x 4o 5c 1o 8c 1o 5c 8o 5c	95
8	TlBa ₂ CuO ₆	1o 5c 5o 5c	-
9	TlBa ₂ CaCu ₂ O ₇	1o 8c 1o 5c 5o 5c	40
10	TlBa ₂ Ca ₂ Cu ₃ O ₉	1o 8c 1o 8c 1o 5c 5o 5c	123
11	Tl ₂ Ba ₂ CuO ₆	1o 5c 5o 5c 5o 1c 5o 5c 5o	95
12	Tl ₂ Ba ₂ CaCu ₂ O ₈	1o 8c 1o 5c 5o 5c 5o 1c 8o 1c 5o 5c 5o 5c	118
13	Pb ₂ Sr ₂ YCu ₃ O ₈	1o 5c 5o 8c 5o 5c 1o 8c	70
14	HgBa ₂ CaCu ₂ O _{8.26}	5o 5c 1o 8c 1o 5c 5o	134

Table 2 Layers sequences of representatives for main classes of perovskite-related materials

least one type of these layers presents in all extracted compounds, that is the main common feature of copper-oxide perovskite related structures. Although the pattern 1' (B) geometrically is similar to 8 (A), it is more closely related to the pattern 1, where all oxygen positions are vacant. Such layers occur in tetragonal-123-O₆ related phases. The another most commonly occurring pattern is 5 (AO), which represents metal-oxygen layers, where the metal usually is an alkaline earth or transition element. Only 2 compounds are found without such types of layers nonetheless they contain the combina-

tion 8o9c (A)_o(O)_c, which is equivalent to 5o and arises from its too large distance between metal and oxygen layers.

Copper-oxide superconducting structures are mostly often described by 3-dimensional primitives: perovskite and intermediate blocks, and coordination polyhedra. They provide convenient structural representations and also play an important role in several theoretical models. At the same time layers have close relations with 3D primitives. The latter can be easily obtained from sequences of layers. Table 3 shows some examples of such relation between layers and 3D primitives in

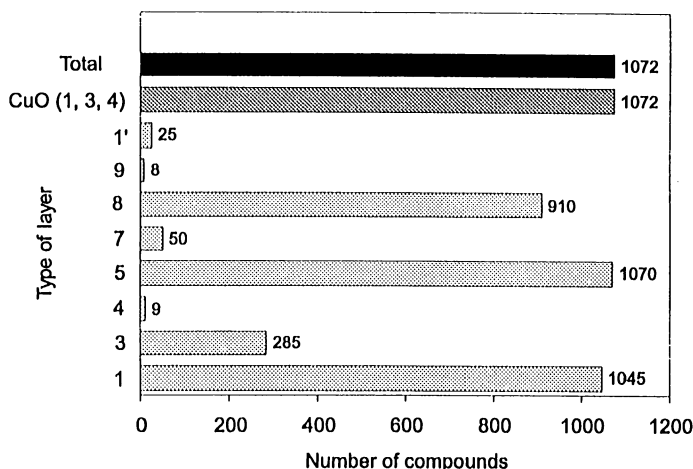


Figure 7 Presence of layers with different patterns in copper-oxide structures extracted from ICSD

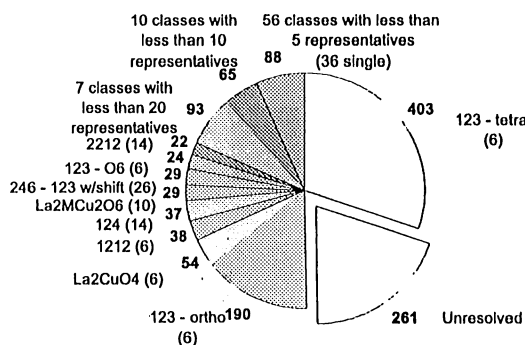


Figure 8 The most representative classes of extracted materials combined by their sequences of layers. Number of representatives is given in bold. For accentuated classes number of layers per unit cell (the sequence length) are shown in brackets.

cuprate structures and emergences of several 3D primitives in the extracted compounds.

All resulting sequences have been sorted as character strings and 82 different classes have been revealed. Diagram on Figure 8 separates the most representative 10 classes (with more than 20 representatives), others are shown in the groups with less than 20, 10 and 5 representatives.

Among compounds in the last group 36

compounds do not have analogs with the same layer sequence. Several reasons can cause this situation:

- Very rare structure.
- Sequences have not been standardized. Structure may be equivalent, but shifted relatively each other in some direction. For example, sequences 1o5c1o8c1o5c1o⁽¹⁾, 1c5o1c8o1c5o1c⁽²⁾ and 8o1c5o1c5o1c8o⁽³⁾. Structure (2) shifted relatively (1) in the layer plane to the center ($x+1/2$, $y+1/2$); whereas sequence (3) shifted in the direction perpendicular to the layer plane. To eliminate separation of these structures, a normalizing procedure is necessary.
- Layers separation due to large constituent elements, as in case of 5o=8o+9c.
- Various kinds of errors in database. Unfortunately, as any man-made product, ICSD is uninsured from some errors, although, all data go through serious procedure of quality check. However, for example, 5 of the extracted compounds have interlayer distances more than 5 Å that seems to be impossible. They are displayed in Figure 9 by cross sign (×) and have been discarded from

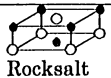





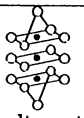
Sequence fragments	5o5c	5o1c5o	8o1c5o	8o1c8o	5o3(4)c5o	5o1?fc5o	1o8c(1o8c) _n 1o
3D primitives	 Rocksalt shift	 Cu-O octahedron, Perovskite block	 Cu-O square pyramid, defective perovskite	 Cu-O square	 Cu-O square chains	 Cu-O chains	 alternate Cu-O layers
Number of compounds	298	614	904	45	208	24	877

Table 3 3D structural primitives as fragments of layers sequence.

the analysis. In several cases a wrongly assigned space group or atomic coordinates cause a distance between two atoms to be significantly smaller than the sum of their ionic radius. Thus the proposed procedure can serve as an additional tool for data quality control.

By the same reason as presented above programs are also imperfect and can result in an error in some unusual situations. To eliminate them wherever possible, sequences have been analyzed by several criteria:

- symmetry of layers relatively center of sequence;
- combinations of neighboring layers following Table 1 (revealed variants are shown there in bold);
- average distances between layers calculated from c parameter of unit cell and number of layers per unit cell and comparing them with Figure 9.

Last criteria, average interlayer distance d_l , seems to be very important for describing and understanding similarities in layered structures. Consequently, such an analysis has been performed in a systematic way and is discussed here.

To know the influence of the number of layers on unit cell size, the dependence between parameter c and number of layers per unit

cell has been illustrated. (Actually the parameter c in this context does not mean the lattice parameter c assigned in the database, but the maximum axis, which represents the direction of layers stacking.) The result is shown in Figure 9. It proves a linear character of their relation for all analyzed structures. Data points can be approximated by a trend line $y=1.95x$ at R-squared value $R^2=0.98$ and under a condition of fixed intersection of the origin of coordinates. (Under this approximation the thickness of layers is not taken into account, thus at $N_l=0$ $c=0$.) Coefficient 1.95 is nothing but the average interlayer distance D_l in angstroms among all analyzing compounds. The upper and lower boundary assuming average interlayer distance (± 0.5 Å are shown as dashed lines. The average interlayer distance is obtained by another way, by averaging ratio $d_l=c/N_l$ for each compounds. In this case D_l equal to 1.96, that is very close to previous result. These data also are consistent with the results obtained by calculation of interatomic distances in the direction perpendicular to the layer plane, although in that case the thickness of layers has been taken into account.

As one of the most important topics on these superconducting and related materials is to find a guideline to improve such properties as T_c , J_c , H_c , it seems to be worthwhile to test whether there is any correlation between structural parameters and critical tem-

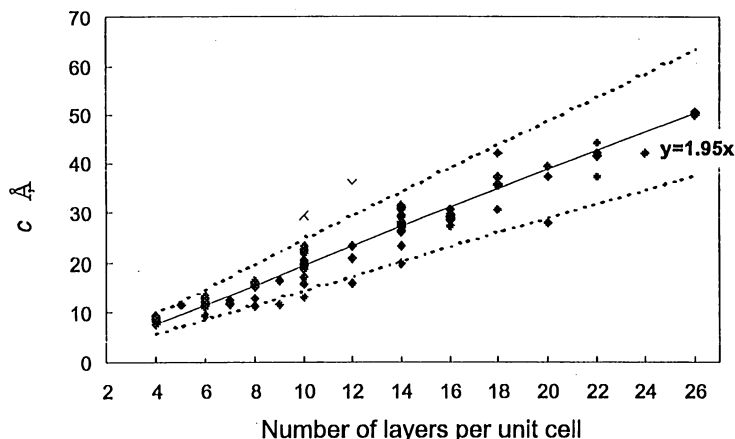


Figure 9 Lattice parameter c via number of layers per unit cell in copper-oxide compounds extracted from ICSD

perature of superconducting transition T_c or not. As a trial example, T_c via average interlayer distance (c/N_l ratio) are displayed in Figure 10.

As has been emphasized before ICSD database does not contain any physical properties including superconducting ones. The T_c data have been obtained from papers cited in ICSD. Therefore, these data has been obtained by various techniques, on different samples, by different researchers and in non-standardized experiment conditions. Thus they can be used in such study only with a great caution.

Unfortunately in this case even without any mathematical treatment the lack of any correlation between these properties can be concluded. It implies we need additional parameters to find some correlation in another direction or possibly in the aggregation with other dimensions. Nevertheless the study of such dependencies and search for reliable structural parameters should be continued. Several perspective-directions can be proposed. For example, thickness of layers (both average and in specific layers), distances between characteristic layers, size of structural blocks and so on. These and other features are planned to be implemented in the future versions of this developed software.

8 Search for compounds with isomorphic structure

There is also a problem now to be able to state any structure as a model and search for isomorphic structures. It is not a trivial task. Similar structures may appear in different space groups or with a different formula making it difficult to prescreen. Relation between isomorphic compounds often can be missed without prior chemical knowledge. On the other hand sometimes it is difficult to separate non-isomorphic structures with similar elemental composition. For example, let's consider 3 compounds from the ICSD:

All of them have very complex elemental compositions with element deficiency and substitution. Is it possible to find some relations between them without 3D representation? Somebody may find that elemental ratio of these phases is very close to 123 ($\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$), compound 4 in Table 3:

1. $\text{PrBa}_2(\text{Cu}_{2.83}\text{Al}_{0.17})\text{O}_{6.55}$
2. $(\text{Ca}_{0.70}\text{La}_{0.30})(\text{BaLa}_{0.56}\text{Ca}_{0.44})$
 $[(\text{Cu}_{0.95}\text{Zn}_{0.05})_2\text{Cu}_{0.78}] \text{O}_{6.52}$
3. $(\text{Y}_{0.63}\text{Ca}_{0.53})\text{Ba}_2(\text{Hg}_{0.63}\text{Cu}_{2.21})\text{O}_{6.6}$

However, layers sequences for these compounds:

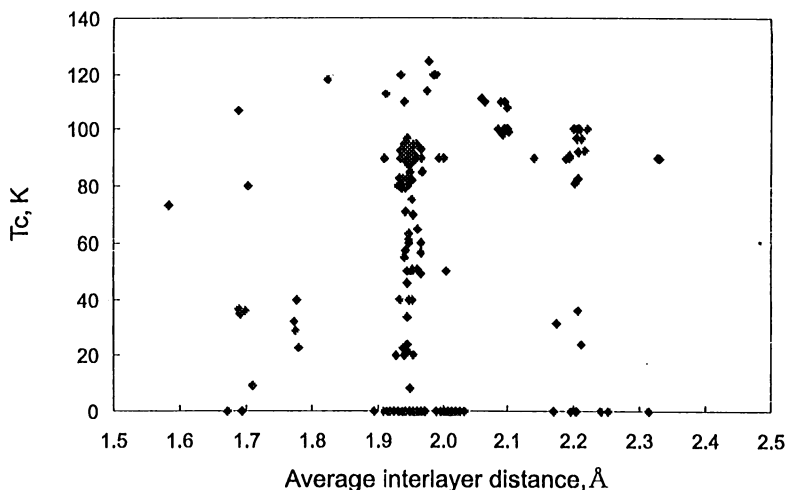


Figure 10 Critical temperature of superconducting transition via average interlayer distance in copper-oxide compounds extracted from ICSD

N	Formula	Space group	a, Å
1	Ba ₂ PrCu ₂ (Cu _{0.83} Al _{0.17})O _{6.55}	P4/mmm	3.906
2	(Cu _{0.95} Zn _{0.05}) ₂ (BaLa _{0.56} Ca _{0.44}) (Ca _{0.70} La _{0.30})Cu _{0.78} O _{6.52}	Pmmm	3.877
3	(Y _{0.63} Ca _{0.53})Ba ₂ Hg _{0.63} Cu _{2.21} O _{6.6}	P4/mmm	3.87

1. 1o 5c 1o 8c 1o 5c 1o
2. 1o 5c 1o 8c 1o 5c 1o
3. 5o 5c 1o 8o 1o 5c 5o

give us more close answer. First two compounds have identical sequences of layers and belong to 123 class, whereas third one is Hg-1212 phase.

It seems possible to conclude that this approach to description of layered compounds is very useful for comparing structures to one another, for analysis of their geometrical relationship and for understanding the atomic configuration of complex structures.

9 Conclusions

In this work the example of definition and isolation of structural primitives for the case of copper-oxide high temperature superconducting and related materials has been de-

scribed. Similar approach can be applied for other classes of materials by revealing off their common structural features, structure primitives and rules of their organization. As was shown the layer-by-layer model help us not only to retrieve useful information about key elements in crystal structures of currently known superconductors, but also to find relations between different structures even with different structural types and space groups.

Acknowledgments

Y. Kotliarov was supported by the scholarship of Japanese Ministry of Education, Science and Culture (Monbusho) during his doctor course.

References

- [1] Shaked H., Keane P.M., Rodriguez J.C., et al., Crystal structures of the high-temperature superconducting copper-oxides, Elsevier, Amsterdam (1994).
- [2] Tokura Y., Arima T., Jap.J.Appl.Phys., 29 (1990), 11, pp. 2388-2402.
- [3] Santoro A., Beech F., Marezio M., Cava R.J., Physica C, 156 (1988), p. 693-700.
- [4] Kovba L.M., Lykova L.N., German M., Antipov E.V., Zh.Obschei Khimii, 56 (1986), 5, p. 1006-1014 (in Russian).
- [5] Cava R.J., Science, 247 (1990), 4943, p. 656-662.
- [6] Santoro A., J. Alloys and compounds, 197 (1993), p. 153-158.
- [7] Poole C.P., Datta T., Farach H.A., J. Superconductivity, 2 (1989), p. 369-386.
- [8] Izumi F., Takayama-Muromachi E., High-temperature superconducting materials science and engineering: new concepts and technology, Pergamon Press, Oxford (1995), chapt. 3.
- [9] Solodovnikov S., Problem of Crystallo-chemical Design of Superconducting Oxides. Preprint of Inst. of Inorganic Chemistry, Novosibirsk (1990), p. 15-22 (in Russian).
- [10] ICSD (Inorganic Crystal Structure Database). User Manual. Feb.1995.
- [11] <http://www.rhrz.uni-bonn.de/~unc442/icsd0.html>
- [12] <http://kristall.erdw.ethz.ch/sginfo/>
- [13] Yvon K., Francois M., Z.Phys.B, 76 (1989), 4, p. 1519-1527.
- [14] Pauling L., J.Am.Chem.Soc., 51 (1929), p. 1010.
- [15] <http://www.pdb.bnl.gov/>
- [16] Zallen R., Proc. of the 12th Intern. Conf. Phys. Semicond. Teubner, Stuttgart (1990), p. 446.

QUEST Doctoral Course, Univ.Tokyo.
Working in JST Corp. in the team of LPF
project from 1996.

岩田修一 (正会員)

東京大学大学院 博士課程修了。
工学博士。

現在、東京大学人工物工学研究センター・
工学系研究科 教授。

卒業より一貫して材料設計のための情報
システムの研究・開発に従事。

材料を主な対象としつつ、人工物全般に
共通する普遍的な原理の抽出と実践的手
法の開発を追及中。

(2000年2月10日受付)

(2000年7月21日採録)

著者紹介

Yuri Kotliarov

Graduated from the Chemical Department
of Novosibirsk State University(Russia).

抄録

大規模ヒト3次元脳画像データベースの構築

志田 和人、川添 良幸、桑川一也

川島 隆太、大槻 昌夫、福田 寛

J.J.S.I.K., Vol.10, No.2, pp.2-11(2000)

今日では、MRI(核磁気共鳴断層撮影装置)などの医用画像機器が高度に発展し、またヒトの脳に関する医学研究上基本的な機器となりつつある。しかし、如何に大量のデータを収集したとしても、それを管理し検索する適当な手段が無くては無意味である。

我々は1000の健常者の3次元脳MRI画像を含むデータベースを構築中であり、これについて報告する。また、この種のデータベースについて要求される一般的事項の幾つかについても議論する。

材料設計のための結晶データマイニングへのアプローチ

ユーリ・コトリヤーロフ、岩田修一

J.J.S.I.K., Vol.10, No.2, pp.12-25(2000)

材料における化学組成、結晶構造、物性それぞれの関係を体系的な分類に基づいて総合的に理解することは、材料設計の基本である。構造や物性値を数多く格納したデータベースは、上記の関係に対する新しい知識を得るための強力な道具となる。ここでは、データベースを利用して無機材料における結晶構造と構造要素データマイニングを行なう上での問題点を、結晶学的見地から述べる。層状銅酸化物超伝導材料における、新たな探索方法の開発と、複雑で未だ発見されていない規則性を見出すことを目的として、二次元積層構造に関する分類を行なった。

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 af-

filiating 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 systematization 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 controlling 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 YAEHASHI 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 infrastructure such as databases. Vast majority of objects with educational, academic, and/or cultural importance are left unexplored. This study aims to establish a basis for the analysis and modeling of information conveyed in these previously unexplored multidisciplinary 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, downloading, 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.

研究部会報告一抄録

情報知識学会 人文・社会科学系部会主催
第12回 歴史研究と電算機利用ワークショップ

人文・社会科学系部会企画担当 田良島 哲

主催：情報知識学会人文・社会科学系部会・学習院大学史学会

期日：1999年11月27日（土）

会場：学習院大学北1号館308号室

「歴史研究と電算機利用ワークショップ」は、歴史的な方法をとる研究分野におけるコンピュータ活用の経験を情報交換するために、情報知識学会人文社会系部会の企画として、継続して開催している。第12回は、神立孝一氏が企画の主担当となり、長谷川順二（学習院大学大学院）・百瀬康司（東京工業大学大学院）の両氏と田良島哲の3名が報告した。また、今回は初めて学習院史学会と共催し、報告者の選定・会の運営を共同して行った。

村松弘一氏をはじめ、学習院史学会のスタッフの方々には当日の会場の設営をはじめ多大なご協力をいただいた。記して感謝申し上げる。

衛星画像を使用した黄河下流域の河道変更
に関する考察

学習院大学大学院 長谷川順二

黄河は古来より幾度となく氾濫し、その度に流路を変化させてきた。今までに数多くの研究者が様々な方法で、その河道を確定させようと試みてきた。史書には都市や他の自然地形との相対的な比較によって河道の位置を間接的に表現しているため、現在に至るまで黄河故河道の正確な位置は確定されず、研究者によってまちまちとなっている。この不明瞭な状態を打破するために、史書研究とは別の資料、特に科学資料を導入しようと思う。

今回私が導入を試みた研究方法是、「リモートセンシング」と呼ばれる。これは中国語では「遥感」と意識されるように、「離れた場所から目的の場所を観察する」観察方法の総称であるが、現在では特に人工衛星によって撮影された画像を使用した研究を指すことが多い。「人工衛星で画像を撮影」といっても、通

常の航空写真のようにカメラを用いるのではなく、センサーによって撮影された画像である。センサーを用いて収集したデータを数値に応じてグレイデータに変換して、地形写真のように加工したのがいわゆる「衛星写真」であって、通常は白黒で表現されている。センサーによって得られるデータには可視光だけでなく、赤外線を感知することもできるので、適切な処理を施せば、通常の航空写真とは異なる結果を得ることも可能である。「画像処理」とは、端的に言えば「画像内にある同様の特徴を持つ範囲を、目に見えるように色分けする」方法である。処理を行う範囲に含まれる地形の特徴によって、様々な画像処理の方法が存在する。

今回行なった方法は、まず最初に目的の画像に画像処理を行なって、河川やそれに類する特徴を持つ地形を強調する。次に処理画像を観察して、河川と思われるラインを手作業で拾い上げる。最後に拾い上げたラインを現

在の地図と重ね、現在の都市や地形を重ね込む。ここまでが画像関連の作業である。

次に正史などの歴史資料との比較を行う。具体的には、先程処理した河川の近辺に所在する都市の記述と、処理画像との位置の相違を比較する。年代によって河道の走る位置が違うので、資料が十分揃えば年代別の河道を確定することも可能であろう。

「リモートセンシング」を歴史学へ応用した例としては、考古学分野において発掘範囲を事前に確定させたことが挙げられる。処理画像を使用することで、ボーリング調査を用いずに遺跡の所在を確定させることが可能となった。また現在残っている都市遺跡を上空から撮影して、当時の町並みや都市計画を推察することも試みられた。いずれも歴史資料を補完する方法として「リモートセンシング」データが活用されており、本研究も同様の研究方法になる。しかし以前の研究が主に乾燥地帯への適用に留まっていたのに対して、本研究では湿潤地帯の河川を研究対象としている。「リモートセンシング」の歴史学への活用の幅を広げることができれば幸いである。

Windowsにおけるマルチリンガル環境

東京工業大学大学院 百瀬康司

1990年代後半は、まさに「インターネットの時代」であった。「インターネット」や「電子メール」が一般家庭に爆発的に普及したことは、同時に、その端末としてのパーソナル・コンピュータ（IBM PC/AT 互換機）と、そのOS（Operating System）であるMicrosoft Windows98/95が普及したことを意味する。

今でも、コンピュータは「計算機」と呼ばれる。たしかに、黎明期におけるコンピュータは、ただの「計算機」であった。だが、現在のコンピュータは、それほど単純な機械ではない。ひと昔前のスーパーコンピュータ以上の性能を持ち、多種多様な処理を行うこと

が可能な、いわば「多目的ツール」へと進化したのである。とはいえ、原型が「計算機」であったことの「弊害」は、いまだに残っている。その端的な例が、文字コードの問題である。コンピュータは、「計算機」として使うことを前提に設計されていたがゆえに、そもそも複数の言語（文字）を処理すること自体、まったく想定されていなかった。そのため、様々な国の企業や規格団体が、自分たちの都合でそれを拡張し、多種多様な文字コードを設定してしまった。このようにして規格化された文字コードの背景には、その国の文化的側面が強く反映されており、一朝一夕に解決できる問題ではない。今後、歴史研究に限らず、あらゆる分野の学問がコンピュータとインターネットを利用した研究方法を導入していくことになるだろう。その際、文字コードの問題は非常に重要な要素となるはずである。とくに、歴史研究の場合は、その性質上、マルチリンガル環境の実現が「必要不可欠」である。決して「ひとごと」ではない。したがって、文字コードとマルチリンガル環境について現時点における問題点を整理し、今後いかなる改善が必要なのかを考えることは、歴史研究に関わる人間にとって義務である、といっても過言ではないだろう。

本発表であつかう問題は、以下の5つである。

- (1) 「文字コード」とは何か？
- (2) 「文字化け」の仕組み
- (3) Unicodeはマルチリンガルへの最短距離か？
- (4) Windowsのマルチリンガル環境
- (5) まとめと今後の展望

なお、発表に際しては、専門的な説明は極力省略し、実例をデモンストレーションすることで、「感覚的」に理解できるような形にする予定である。

といった実践的な課題について、若干の提起を行う。

古典籍のテキスト空間と論理構造—電子化の前提としての検討—

文化庁文化財保護部 田良島 哲

現在、コンピュータとソフトウェアの能力の進歩はいちじるしいものがあるから、印刷で用いられている組版のスタイルを、形の上でディスプレイ上に表示することは可能であろう。しかし、古典籍テキストの電子化で、本質的な課題は「印刷されたとおりに表示すること」ではない。テキストの電子化の上で重要なのは、さまざまなスタイルで配置された文字列群の平面空間内における意味あいと、文字列どうしの関係である。いいかえれば、空間的な文字列群の配置を、論理的な構造に変換することが重要なのである。この点は、意外にも、文学研究者や歴史研究者にも十分理解されていないように思う。

本報告では、まず、漢字を主体としたテキストの空間的な配置が、どのように進化してきたかという点を、検討する。中国において、金石文にはじまり紙に定着するテキストの構造は、その歴史の中で、割書・傍書・頭書など付加情報を平面空間内に記述する方法が開発され、11世紀には、これらの方法を集大成した宋版という東アジア世界に大きな影響を与える組版形式が出現し、一つの規範となる。このような規範は7世紀以降、日本に順次移植され、文書や典籍の表現に用いられるようになるが、訓点のように文字の一つ一つをも空間として使う独自の手法も発明されている。このような歴史的な前提を念頭におきながら、種々の記述形式について、その本文に対する論理的な意味づけを確認してゆく。

その上で、古典籍テキストをXMLのようなマークアップ言語の形で構造化するためにどのような点に留意すべきか、また実際に電子的に編集を行うために必要なツールに求められる条件としては、どのような点があるか、

研究部会報告一抄録

情報知識学会 人文・社会科学系部会主催
第13回 歴史研究と電算機利用ワークショップ

人文・社会科学系部会企画担当 田良島 哲

期日: 2000年5月13日(土)

会場: 法政大学ボアソナードタワー9階 サイエンスルーム

「歴史研究と電算機利用ワークショップ」は、歴史的な方法をとる研究分野におけるコンピュータ活用の経験を情報交換するために、情報知識学会人文社会系部会の企画として、継続して開催している。今回は、独特の記述システムを持つ古典籍や古文書を電子化する上での新しい試みについて、谷本玲大（茨城大学）、鈴木卓治（国立歴史民俗博物館）、小口雅史（法政大学）、小畑真帆（株）フラックストーン）の4氏から3本の報告をいただいた。いずれもXML、PDF、LaTeXなどを駆使した実践的な内容であると同時に、今後のテキストの電子化に関する理論的な問題も示唆するもので、報告終了後、熱心な討論が行われた。

なお、会場の準備と当日の運営については、小口雅史氏に多大なご協力をいただいた。記して感謝申し上げる。

XMLによるタグ付けを用いた古典籍の字彙DBについて

一伝本の性質を把握する補助手段として一

茨城大学 谷本玲大

本発表では、古典文学研究を主とする人文系研究の立場から、XML文書+今昔文字鏡TrueTypeフォント+JavaScript+html出力という組み合わせで作成可能なXMLによる変体仮名字母データベースの紹介を行うものである。

今回発表するXMLによる字母データベースは、コンテンツ内容として国宝、尊経閣文庫本『土左日記』を採用したが、コンテンツ素材が何であれ、同様の処理が可能のようにDTDを設計してある。従って、タグ付けの煩雑さを厭わなければ、比較的容易に変体仮名の実際の字形表示と、それを包摂/区別した検索とを行うことができるものである。

文学研究におけるコンピュータ利用について、本文研究の第一歩である文献学的研究に

応用しようとする試みは、既に統計学的な見地から導入されてきた（行動計量学シリーズ6『真贋の科学—計量文献学入門—』村上征勝著 朝倉書店 1994）。

しかし、国文学に於いて、古典本文は一般に現代では用いられることのない様々な字形を複雑に使用して表記するので、いわゆるJIS漢字以外の多種多様な文字を区分して操作できるようにする必要がある。

だが、DB構築に際し、使用文字をJIS漢字に限定し、IDを付加するのみでは視認性に欠け、DBそのものが着眼・発想を支援する役割を担うことはできない場合が多いと思われる。

故に、コンピュータを国文学研究の道具として使えるようにするためには、様々な字形を表現し、かつ、ある程度は原本の改行や字の割り付け組版を含む表示をする機能が用意される事が本来的には望ましい。

更に、様々な変体仮名字形を詳細に弁別するのみでなく、それらの変体仮名を例えば「お」

という読み方をするもののみ一括で抽出したり、あるいはその「お」という読み方をする文字の中で「於」という漢字がその変体仮名の元となる字母であるもの…など、様々なレイヤーを持たせて検索することができるようになっている必要がある。従来は、これらを特定の読み方をする仮名のレベルや、特定の字母のレベルに丸め操作を行って研究されてきたわけだが、SGML や XML を用いれば、視認性を維持しつつ丸め操作／詳細な弁別操作を様々に行えるようになると思われる。

そこで、今回発表者が紹介する方法は、それほど特殊なソフトウェアを必要とせず、パーソナルコンピュータ環境で簡便に行えるという視点から、XML を採用してコンテンツ構築を行った。

本発表で実演するコンテンツは、XML というマークアップ言語本来が持たされているマークアップ言語としての用途や設計思想からは乖離した利用法ではあるが、XML の利用によって文献学的研究（あるいは、国語学的研究にも転用可能）に着眼・発想を支援するためのDBとして、かなりの要求を満たし得ると思われる。XML によって得られる恩恵と共に、今後の展望、またあわせて実作業に於いて遭遇した問題点などについても提示してみたい。

なお、本発表の元になったのは 1999 年 9 月 16 日から 10 月 8 日にかけて東京外国語大学アジア・アフリカ言語文化研究所短期共同研究員として、同研究所教授芝野耕司氏の指導のもと行った「XML によるタグ付けを用いた『新撰萬葉集』伝本の数量的比較研究」と題する共同研究である。

LaTeX による古文書組版の一経験

国立歴史民俗博物館情報資料研究部

鈴木 卓治

筆者は、博物館における情報システムに関する研究を業務とする情報系研究者である。

古文書はまったくの素人であるが、1998 年 10 月 20 日から 12 月 6 日にかけて開催された本館企画展示「収集家 100 年の軌跡— 水木コレクションのすべて—」の出し物のひとつとして、水木家資料中世文書 154 点のマイクロ写真とその釈文を並べて表示するコンテンツ (http://www.rekihaku.ac.jp/gallery/mizuki/index_mizuki.html にて公開中) を開発した際、LaTeX を利用した古文書釈文の組版を経験した。本発表は、このときの経験をベースに、pLaTeX2e+ 金水マクロ + 今昔文字鏡 TrueType フォント + bookworm マクロ + Acrobat という組み合わせで、海外の環境でも TrueType 品質で表示印刷できる釈文データの作成を実演するものである。

歴史学におけるコンピュータ利用を考えると、その第一の基礎をなす対象は、かたや古文書であり、かたや屏風・絵巻・浮世絵などの図像情報であろう。古文書は一般に複雑な表記規則をもち、いわゆる JIS 漢字以外の多種多様な文字を含み、組版の位置関係や加筆・訂正・削除その他の編集痕といった重要な付帯情報を含んでいることなど、情報工学の観点からみても、その扱いには研究が必要であり興味深い対象である。

コンピュータを古文書研究の道具として使えるようにするためには、組版を含む表示・印刷の機能、入力・訂正・切り貼りなどの編集（しばしば組版指示作業と同時進行）の機能、および検索の機能が用意されなければならない。現時点での最適解は、Microsoft Word などのプログラミング可能なワードプロセッサシステムをベースに、各機能を満たす道具を開発することであろう。

今回筆者が紹介する方法は、対象を組版に特化していることと、組版システム自体が人文系一般になじみのうすいものであることから、広くおすすめてできる方法とはいえない。しかしながら、現在望める最高品質に近い組版データが比較的安価に作成できることや、

指示したい組版情報が確定している場合、組版システムを利用したほうが、WYSIWYGの原則に基づくワードプロセッサよりもかえって少ない手間で組版の指示が行なえることなど、押さえておくべきいくつかの重要な原則を含んでいるように思われる。

日本古代文書の電子化をめぐる諸問題 ーデジタル古文書集作成の実体験をふまえてー

法政大学第一教養部 小口 雅史
(株)フラックス・トーン 小畑 真帆

本報告は『デジタル古文書集日本古代土地経営関係史料集成』（同成社刊）の製作経験をふまえて、現状の問題点と今後への展望を考えるものである。

日本古代の代表的古文書群である正倉院文書は、二次利用されたために残ったという特殊な経緯を有する。また江戸時代後期の整理の過程で、二次利用面が、不用意に截断され一次利用面への回復が試みられたため、現状は、一次利用面も二次利用面も、ともに当時の生の状態ではなくなっている場合が多い。こうした接続の問題と、さらに数次わたって書き込みがなされているという重層性によって、この一大古文書群を詳細に分析するためには、紙媒体の翻刻よりは電子媒体による翻刻こそが、研究上、大きな効果を持つ。

多数のJ I S外漢字を有する古文書群であるが、その問題は、今昔文字鏡によってほぼ解決のめどがついている。縦書き表示はPDFで処理すれば問題ない。残された課題は、ダイナミックな検索プログラムの開発である。

今回は以下のようなステップを踏んだ。

○古文書のデータ処理

- ・原典図版を座標系にマップする
- ・単語区分の識別
- ・検索アルゴリズムの重層化

この過程で、日本古代文書の電子化をめぐ

る諸問題がいくつかうかびあがった。必要なのは(既存のシステムの拡張も含め)、古文書ページ記述システムなのだろうか。主要な開発環境すべてに標準プログラムソースをライブラリとして用意すべきなのか。標準辞書として、完全な辞書には割注や文字サイズに関する情報も必要で、さらに、その影響力の調整機能は実現してみたい。ではオブジェクトデータベースとして理想的な結論はどこにあるのか。今回は実践例をふまえての諸問題の提示に留まらざるを得ないが、当日の議論の展開に期待するところも大きい。

XML 解説

新たな生産性原理とドキュメント・マネジメント

- XML ドキュメント・データベース・システムによるリエンジニアリングの達成へ向けて -

(株)ドキュメント・エンジニアリング研究所 西村 健

1 IT 革命とは

本年に至り「IT 革命」が流行コトバとして急激にひろまっている。1990 年代の我が国では、「情報化」や「コンピュータ・システム化」が前向きな話題として迎えられてはいても、IT 革命すなわち「IT による社会変革」という広範かつ根源的な影響力の発現を意味するキーワードに関しては、過激な表現であるとして批判的な受け止め方が大勢であったことを省みれば、かようなコトバの急展開ぶりには、本質と向き合うことを避け、手っ取り早く流行に乗ろうとする怪薄さをいつもながらに感じるところである。

ひるがえって世界を見れば、「CALS」「情報スーパーハイウェイ構想」「リエンジニアリング」などのキーワードにより周知のごとく、1990 年前後から IT による社会変革が先進的国々において真正面から広範に論じられ、政策の中枢に据えられるとともに、社会システムの各分野を通じた変革の梃子としての位置付けと具体的取組みには、徹底したものが認められる。

もとより「変革」とは、新たなフレーム・ワークの樹立に始まり、新たなシステムの設計、構築とともに既存システムの解体を伴うシステム転換が不可避で、莫大なエネルギー投入によりはじめて成り立つ営為である。こうしたことから、大局的な全く新たな枠組みを打ち立て、そこを目指して大幅な変更を強いることとなるダイナミックな戦略的アプローチ法は、仮説設定に基づくリスクな方法であり、一般に「変革」より「カイゼン」をよしとする日本社会にあっては、できるだけ避けたいとの気持ちが根強く存在するかのようである。

しかるに、グローバルな競争社会が広がりつつある今日では、否が応でも生産性のあり方自体を問い直し、大胆に変革できなければならない状況に突入しているといえよう。このような日本社会にあって、IT 主導による生

産性原理のパラダイム・シフトが不可避とされる本質的理由、及びそれを実現する考え方と道筋とはどのようなものか。

この課題をめぐり本稿では、過去数十年の間に主として我が国とアメリカで繰り広げられた生産性向上へのシステム転換について若干の考察を加え、今日求められている新たな生産性原理の観点からリエンジニアリングの意義を再確認する。それを元に、附加価値生産性を支える「ドキュメント・マネジメント」の重要性を論じたい。

2 リエンジニアリングの本質

2.1 ホワイトカラーの生産性とリエンジニアリング

これまでの日本にあって、奇跡と評されるレベルの目覚ましい生産性の向上を現出し得た歴史を幾度か有している。そうした場面を例にとってみれば、そこには常に顕著な革新的システム転換が存在していたからこそ、その後の発展が成し遂げられた、と考えるべきであろう。

この観点から、戦後日本社会の生産性向上の主たる牽引力となった「QC 活動」について考察してみると、工場における製造ラインという純然たるブルーカラー業務について、より良い生産のあり方とはなにか、という課題を作業員自らに課する姿となっており、本来は生産管理に従事するホワイトカラーの職分を現場労働者が担うという意味において、「ブルー業務のホワイト化」したもの、と評することができよう。生産ライン自らが品質向上への意欲と思考能力を自律的に保持した、ということである。近代産業社会の典型たる工場生産において、徹底的な効率的生産計画に基づき、分担機能発揮を求められているブルーの生産ラインが、その当初機能設計に反して「思考する生産ライン」となったところ

にエポックが起こった。すなわち革新的システム転換が成立していた、といえるであろう。

ある地域においてこうしたシステム転換が奏効し、生産性向上が製品に反映されると、国際的な競争にさらされる市場では、ただちに抜きん出た競争力を獲得し始め、システム転換非達成地域の製品を駆逐することとなっていく。そこで、新たなシステム転換をめざしたトライアルがスタートした。1980年代、アメリカの製造業は市場での競争力低下という現実直面し、再生へ向けたテイクオフを図る手だてを求めて、日本企業の経営手法の実態をつぶさに調査・研究することにより、QC活動を主とする「思考する生産ライン」の強みを理解するに至った。さらに新たな生産性原理を模索する中から「TQM」メソッドへと拡張・発展させていった。いわゆる「デミング博士の15ポイント」メソッドがこの代表である。

アメリカではやがて、急速にシェイプ・アップしてきた情報技術を積極的に導入し、こうしたメソッドを「エンジニアリング・システム」として発展させ、ついには劇的な生産性向上のシステム転換を成し遂げるに至った。リエンジニアリング(BPR)である。

それではリエンジニアリング(BPR)のエポックたる所以の本質はなにか。まず第一に、生産性向上の狙いとする範囲が全組織を包含した総体性をもつ点である。QC活動では、生産ラインを中心とする製造部門における生産性向上を追及したものであったが、TQMでは、全社、全組織を通じた生産性向上へと拡張されている。すなわち、日本では未踏の領域であったホワイトカラーの生産性を狙上に上げた、という意義が確認されなければならない。

この考え方を土台にしたとき、これまで手が着けられなかったホワイトカラー業務の内実にはじめてメスが入り、ここに、経営組織体は総体としての生産性を論じる下地が形成される。

近代社会の整然たる役割分担組織は、業務プロセスにおける情報・知識の受け渡しにおける整合性と精密さが身上で、これを担保するため、中間的管理方式は肥大化し、煩雑化し続ける。ところが、20世紀末の高度情報化社会では、サービスの附加価値性が重要視さ

れ、提供されるサービスの質に顧客(究極的には個客)が満足できるかどうか問われることとなり、実際にサービス提供される現場(本稿において「CS現場」という)における情報が決定的に重要性を持つ。CS現場と業務プロセス各段階との間における情報・知識の相互活用・連携が齟齬をきたしたり、大幅に遅延し活用不可となれば、決定的に生産性を阻害してしまうからである。そこで、これを避けるため、情報流が現場をキーポイントとして組立てられる必要がある。それには、サービス提供の現場における思考を成立させるだけの情報・知識の集約・関連付けといった知的糾合性能を担保するため、情報技術の成果を結集して時間、空間の制約を取り払い、中間処理や中間介入を排除した「コンカレント方式」をシステム化するというモデルが構成されたのである。

これを要するにリエンジニアリングの本質性として第二に、「中間処理業務の徹底縮減」があげられよう。従来の近代型産業組織にあっては、分業構造の中でホワイトカラーが業務の大部分の時間を投入し、こなしていた情報・知識の受け取り、集約、整理、伝達などの中間処理作業を大幅に縮減する、ということである。

2.2 リエンジニアリングの実効性

このことは、今日の革新的システム転換が、IT主導によりはじめて達成されることを明らかにしている。なぜなら、これまでの経営組織体における業務プロセスでは、すべて物理的な紙文書を媒体として情報・知識が伝達・処理されていた。そこでは、業務プロセスの1段階ごとに紙文書を受け渡すことにより、順番に手続を進める「リニアな業務手順」が厳然としてあり、時間・空間の物理的制約から自由になることは不可能であった。ドキュメント内容を紙媒体から電子媒体へ転換することによって、時間・空間の物理的制約から離れ、複数の関係主体が同時進行的に業務目的へ向けたコラボレーションを成立させ、さらにCS現場との素早い情報流の交換から、生産プロセスにおける最適化へとフィード・バックをかけることが可能となる。

現実に我国の大規模な組織におけるホワイトカラーが従事している業務実態のなかにあつて、こうした「中間処理作業」が大きな阻害要因となっており、従つて、こうした中間処理作業をシステム統合化することにより劇的な生産性向上が達成できる、ということに関して、すでに明晰な実証事例が存在する。

大企業本社の企画・管理部門におけるホワイトカラー業務について、詳細な業務分析を行った上で、それぞれ個別作業のインプット、アウトプットが紙への入出力形態をとっている業務(伝票処理などのいわゆるトランザクション処理業務で、「起票、転記、決裁、機密保持、保管、参照、照合、検証、計算、配信、集信、ソート、削除、廃止、差し戻し」の15形態)について、紙への出力をすべて廃止し、情報ネットワーク上で編集処理できる統合的システム化の構築事例があり、対象となったホワイトカラー業務の工数が7割以上削減されたというものである^[4]。

ただ、この業務分析によれば、対象とされたいわゆるトランザクション処理業務の占める割合は、ホワイトカラー業務全体の3割程度であり、残りの6割余りは「創造思考的作業」ないしは「準思考的作業」という分野と認識され、システム化の対象外であった。そこで、残された課題は、一般的ホワイトカラー業務の過半を占めると想定される、こうした「創造思考的作業」ないしは「準思考的作業」と名付けられている業務分野について、リエンジニアリングの対象に取込むことである。

2.3 附加価値生産プロセスとしてのリエンジニアリング

「創造思考的作業」ないしは「準思考的作業」とは、どのような特性と作業内容をもつか。まず事柄の性質上、あるテーマに関連する情報・知識の収集・整理が基本とされる。すなわち、ドキュメント内容の意味の読み取りにより、必要箇所を取捨選択し、さらに関連する情報・知識を他のドキュメントから収集・整理する、という行為の繰り返しから、徐々に「煮詰め」「幅狭め」を行い、課題のソリューションに接近していくプロセスである。こうした試行錯誤による知的生産の方法では、「KJ法」の図解展開法に見られるごと

く、関連する情報・知識を近くに配置し、全体として新たな意味のある構造として認識できるレベルへと一定の秩序ある取り纏めを行うまでに、膨大な時間を投入することとなる。

情報・知識に一定の編集処理をほどこし、判断を行うタキ台として提供する、という行為は、「段取り八分」と表現されるようにそこへ至る前段の整理作業に没入することが、すなわちホワイトカラーの生産的活動そのものと見られてきた。しかし、すこし考えてみれば、この作業もあくまで「作業」であるに止まり、そのこと自体からは積極的附加価値を産み出さない。従つて、先にリエンジニアリング対象としたトランザクション処理業務におけると同様に、「段取り」のための編集処理作業は、システム化により大幅な工数削減がなされねばならない。

そこでこの編集処理作業に関し考察をまとめてみれば、次のようである。

ホワイトカラーが従事する業務の中で、ある文書からその内容を理解し、そこから得られた意味を活用してひとつのアウトプットを生みだそうとするとき、たんに当該文書だけを読み解くことをもって足りる、という場合はごく稀であろう。多くの場合、複合的な関係性を持つ文書群が存在しており、業務の遂行に当っては、それら文書群の関係性からもたらされる意味理解に立脚することが要求されている。こうした意味理解に達するには、ある文書と関連する他の文書の当該関連部分を、一定の処理方式を加えることにより関連付け、一つながりのコンテキスト(文脈)としての意味を汲み取りうる状態にまで、内容を成熟させ高めることが必要である。

「一定の処理方式」とは、「比較、対照、並べ替え、読み替え」などの広義の編集をさすが、こうした編集処理を行うには、簡単な例では関連するページを開いて、突き合わせて読み取る、といった程度で済ませることができようが、やや複雑になってくると、キリバリ加工したり、ワープロで一覧表にまとめ直したり、という手間が掛けられてきた。一般に資料づくりといわれる業務作業の多くがこうした内容を持っており、様々な組織でホワイトカラーが行っている業務の実態でもある。

これらの編集処理について観察してみると、次の3態様の要素を前提として成り立ってい

る。

(1) 複数文書間の関連性の認識

(2) 関連性の態様の解明

- ・参照する
- ・転記する
- ・集める
- ・並べかえる
- ・組合せる
- ・比較・対照する
- ・読み替える
- ・ルーチン进行处理する
- ・その他

(3) 関連性を担保する識別記号の発見又は付記

すなわち、関連箇所どうしを識別できる記号を文書の中にあらかじめ見付けておく、ないしは、何らかの識別記号を付与しておく。その識別記号をキーとして、これら複数文書間の関連性の態様に従って、編集処理を行う、ということになる。

以上の考察から、関連する複数ドキュメントの内容の特定部分を、関連付けるとともに、当該関連性を部品化して蓄積及び再利用に供することを前提とし、次いで、関連性の態様に応じて一定の編集処理機能を持たせるようにすること、というシステム要件が提示される。

例えば、マニュアルが改訂されたとき、まず第一に、改訂がなされた当該個所の新旧対照表を自動的に編集処理し、利用者の誰にも改訂箇所と改訂の意義が明らかに示せるようにするとともに、第二に、当該マニュアルを引用している他の規定文書について、連動して改訂をなすべきかどうか参照して文脈を確認する、といった具体的場面が考えられる。

現状では、こうした業務場面は、ことごとく紙文書の人力処理に頼っているため、次のような困った事例が頻繁に発生することとなる。

- ・業務と文書の関係など異動時の引継ぎが不完全である
- ・マニュアルが改訂されたが、新旧対照表がなく、どこがどのように改正されたのかわからない
- ・マニュアルは、使用現場により少しづ

つ差異があるため、幾通りもあり、改訂のとき幾通りもの改訂作業をしなければならず、混乱する

- ・申請に対する処理事務の際、関連する必須確認事項をチェックし忘れ、ミスが発生してしまった
- ・〇〇検討プロジェクトの出席者が毎回異なり、従前の検討状況が引き継がれず、足踏みしてしまう
- ・原稿が修正されるたびにウェブへの登載に手間取る
- ・事業説明の調書を作成したが、少し視点の異なる資料調製の要求があるたびに、全部作り直しになってしまう
- ・事業調書を作成したが、上位計画の柱立てが変更され、事業調書を大幅に組み直さねばならなくなった

3 ドキュメント・マネジメントの必要性和 XML による解決

以上の考察から明らかとなったように、附加価値生産を担保する新たな仕組みは、複合的なドキュメントの意味内容を関連付けて取扱うことが前提となる。先の考察に基づくシステム要件に対応できるのは、構造化ドキュメント手法たる「SGML/XML ドキュメント・システム」だけであり、ウェブ上での利用を考えると、ただ「XML ドキュメント・システム」のみがホワイトカラー業務の煩雑な段取り作業を縮減し、附加価値生産活動へ転換できるシステム化への道筋ということとなる。

こうしたドキュメント内容の活用による附加価値生産活動を「ドキュメント・マネジメント」と名づけることとするが、ドキュメント・マネジメントにあつては、いずれの IT を使うべきか、ではなく、まず中核に「XML エンパイロメント」ありき、から出発し、その周辺を取り囲むエンジンやインターフェースとして、どのような IT バリエーションが組み合わされるべきか、というアプローチがとられなければならないといえよう^[2]。

このことはまた、今日のホワイトカラーが IT を活用する場面からみるなら、次のよう

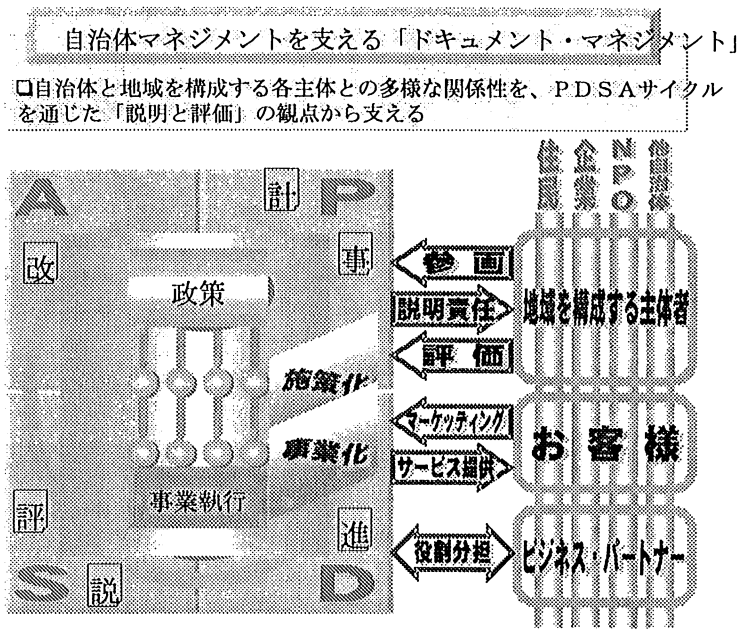


図 1: 自治体マネジメントを支えるドキュメント・マネジメント

に考察される。従来の情報システムでは、集中処理の発想からシステム設計がなされてきた。つまり、データの切り出し方が、利用者にとってカスタマイズできないのが通例であった。ところが、今やCS現場において、エンド・ユーザーとしてのホワイトカラーが附加価値生産活動を行なうとなれば、業務現場の特性に合わせ、各担当者が他と差異化された価値を生み出すために、情報知識の切り出し方、フィルタリング、フォーカスの掛け方が独自の仕掛けを要求することとなる。こうした多面的な活用に対応するドキュメントの内容の柔軟さをXMLエンバイロメントに求めている、と。

4 自治体による取組の重要性

4.1 主体的地域経営とドキュメント

ここまでは、一般企業の経営組織体におけるホワイトカラー業務を例にとり考察を進めてきたが、このことは、自治体をはじめとする公共的経営組織体にあっても、そのまま当てはまる。特に自治体にあっては、地域の生活現場というCS現場における付加価値を目

指す生産活動そのものと考えられる。

日本社会はいま、明治以後の長きにわたる中央集権体制を脱し、自律分権型国家構造へのシフトを求められている。そこで自治体は、地域現場における多様な構成主体間の多様な関係性の中で、「説明責任—行政評価」を手法とする21世紀の分権型社会構造を展開する主体的機能を力強く発揮すべき責務(図1参照)を有していることから、地域現場でのIT革命という新たな生産性原理の樹立と推進を牽引し、成功させるという極めて大きな使命を課せられ、早急な具現化を強く迫られている。

こうした観点からすれば、これまでの自治体業務を担保してきたいわゆる文書管理のライフサイクルの発想は、物理的紙文書のパラダイムからする根本的欠陥を持っている。決裁プロセスを経ていわゆる公文書となったドキュメントとそれ以外の多様な資料の取扱い方の差異、公文書を改竄されないよう固定化した保管方式、一定年限での画一的廃棄処分などは、新たな附加価値生産システムに対応した新たな業務プロセスの担保手段としてのドキュメント・マネジメントへと早急に体制整備、移行を求められる。

4.2 新たな業務モデルとの一体的システム化

このこととも関連して、ドキュメント・マネジメントが実効性を持つためには、新たな組織機能と一体化した業務モデルの創造が伴わなければならない。すなわち情報ネットワーク上で、デジタル・ドキュメントを活用した新たな仕事のあり方を構築しなければならない。2004年に供用開始される「建設CALS」におけるドキュメントの標準化とドキュメント内容の統合的活用へ向けた壮大な取組みが、まさにそうした新たな業務モデルの創造と一体的に進められている。ただ、建設CALSの壮大さにばかり目を奪われていては、本質を見誤ることとなる。最も肝心なことは、リエンジニアリングの定義に関し述べたように、「CS現場に情報流を結集して現場の課題解決に活用できること」が実現されることで、このため、現場のエンドユーザ自身による業務モデルの構築と一体的なドキュメント・マネジメントのリテラシーを獲得することがきわめて重要とされる。

4.3 ドキュメント・マネジメント・リテラシーの獲得へ向けて

筆者らが調査設計し本年実施された「自治体情報化アンケート調査」^[3]の結果によれば、自治体の情報化の現状と課題に関し、次のような点が特筆される。

- [1] 情報化環境整備面の進展と活用実体との乖離
 - ・パソコンの配備、LANの敷設、ホームページ開設は進んだが、その活用のされ方がワープロ、表計算中心で、LAN活用に至らず
 - ・ホームページ開設8割。多様な受発信機能が課題
- [2] 情報化人材不足深刻
 - ・情報化の阻害要因「財政難」トップだがキーパーソンの不在がより深刻
 - ・情報化コア人材の研修実施22%でそのうち8割が年間5時間以内
 - ・特に小規模自治体で情報化人材の不足が深刻

[3] 情報公開と電子文書化の認識乖離

- ・「行政情報化の果たすべき役割」の第一は「情報公開」だが、情報公開の電子化は「未定」が74%
- ・電子文書管理システムと情報公開との関連性認識弱い
- ・電子文書管理システムへのニーズは高いが、まだインデックス系が中心

総括すると、行政改革や情報公開を進める上で行政情報化が必要との認識は浸透しているが、実際の情報化内容はまだまだといえ、特に情報化をめぐる矛盾や問題点として、電子文書管理への取組みが弱く、しかもドキュメント・マネジメントの発想からほど遠いこと、さらに、情報化推進の中核人材育成へのインセンティブが極めて弱いことが明らかとなった。要するに、本論で展開してきた新たな付加価値生産の仕組みを支えるドキュメント・マネジメントにたいする課題認識や取組みがほとんど認められず、これでは何年たってもIT革命の実態に近づくことは不可能と考えられる。

日本社会全体をあげた早急なドキュメント・マネジメント・リテラシー向上策が喫緊の課題と認識するものである。

参考文献

- [1] 日経コンピュータ.1994-11-14 pp.88-96
- [2] 長村玄. XMLによって変わる文書処理のパラダイム. 行政&ADP. 2000年3月-7月
- [3] 自治体情報化アンケート報告書, 社団法人日本能率協会(2000年3月実施.)

書評

計量情報学—図書館／言語研究への応用

影浦峯著 著. 丸善. 2000 年. 182p. 2,600 円(本体)

駿河台大学文化情報学部 岸田 和明

本書で意味するところの「計量情報学」とは、一般には、計量書誌学と呼ばれている図書館情報学の下位領域が発展したものと見なすことができる。この計量書誌学が探究する基本的な問題としては、いわゆる「集中と分散の現象」や「引用分析を用いた学問領域の構造の析出」を挙げることができる。このうち前者に関しては、単語の使用頻度や研究者の発表論文数が具体的な問題として取り上げられ、さまざまな角度から研究が進められてきた。その結果、ジップの法則やロトカの法則と呼ばれる一種の規則性が普遍的に観察されることが知られており、これらの法則のさまざまな性質等が明らかにされている。

本書はこのような計量書誌学的な法則について知るための格好の学術書である。特に、統計的な理論面からきちんと押さえたいという人には最適であろう。さらに、本書の大きな特徴は、計量情報学をこのような計量書誌学と「計量言語学」とを幅広く含むものとして捉え、計量言語学における成果を大幅に取り入れている点にある。この点、本書を計量言語学の問題を知るための本として使うこともできる。著者の影浦氏は、現在、国立情報学研究所で助教授として勤務されており、図書館情報学と深く関わりながらも、これまで、専門用語をはじめとする言語学の問題に関する研究成果を数多く出されてきた(本学会でも何度か研究報告をされている)。このあたりのことが本書にも色濃く反映されているようである。

具体的には、本書では、計量情報学的なデータから算出される統計量の「標本量依存性」の問題に数多くのページが割かれている。これは、計量情報学の基本的な問題であり、計量情報学が扱うデータが、標本平均の分布に

対して正規分布を仮定できるような通常の統計学的データではないことに起因する問題である。誤解を恐れずに簡単に説明すれば、標本のサイズを大きくしていくと系統的に統計量に変化していくという性質こそが標本量依存性であり、この結果、そのような現象を実証的に分析するには、通常の統計学にはない、いくつかの工夫あるいは解釈が必要になる。本書では、まずこの問題の導入に始まり、その具体例、その解釈をめぐる議論、そのような性質を持つデータに対する分析方法などが丁寧に解説されていく。そして最後には、その分析方法の1つである Zipf 族 LNRE (Large Number of Rare Events) モデルの説明で締めくくられている。

このことからわかるように、本書は計量情報学の諸問題を幅広く概観したものではなく、最近の話題の1つに焦点を当て、それを基本から高度なレベルの研究成果まで、根気よく解説した学術書である。そのため、数式も数多いが、その説明・展開等は非常に親切である。類書の中には、数式の説明を簡便に済ませようとして、途中の展開や添え字を省略したために、かえってわかりにくくなっているものが見受けられるが、本書にはそのようなことはない。記号の説明も十分で、何の記号かが明瞭でないためのフラストレーションもおこらない。

このようなレベルの計量情報学(あるいは計量言語学)の単行書を日本語で読めるというのは幸せなことである。この点、著者と出版社とに敬意を表したい。

用語解説

ナレッジ・マネジメント

(株) ジー・サーチ Dialog サービス事業部 長塚 隆

ナレッジ・マネジメント (knowledge management) は新しい経営の指針として、あるいは多くの企業での実践例を通じて最近広く知られようになってきた。しかし、ナレッジ・マネジメントを簡潔にかつ厳密に定義しようとすると、簡単ではない。それはナレッジ・マネジメントが、組織の構成員が有する「知識」を組織全体の活動に具現化するための様々な要因・要素の分析や具体的な行動のすべてを含む概念として考えられることが多いためである。

ナレッジ・マネジメントに関する文献・レポート数の推移を、Dialog の各種データベースで調べた。すると、文献・レポートは1970年代から少数あり、1980年代を通じ徐々に増加するが、その数はそれほど多くなかった。しかし、1990年代の半ばから文献・レポートは急増し、現在に至っている。このように、ナレッジ・マネジメントという用語はすでに1970年代には技術社会の管理に必要な考え方として使用されており、また、1980年代には組織内で個々に作成されたレポートなどの知識を統合的に有効活用するためのソフトウェアとしてのナレッジ・マネジメント・システム (KMS) の考え方として提案されていた。しかし、文献・レポート数は少なく、社会的にはそれほど注目されてはいなかった。

ところが、1990年代半ばになると、インターネット・イントラネットの普及を背景に、文献・レポートが急増し、社会的な関心を集めるようになった。これは、組織内における知識を有効活用するためのシステム構築が容易になったことを背景にして、多くの企業で様々な種類の実践が行われ、多くの文献・レポートが発表されるようになったことによ

るところが大きい。

ナレッジ・マネジメントの定義はこのような状況を反映して、現在、様々な角度よりなされている。その一例を挙げると、「組織内に蓄積された知識をよりアクセスし易くかつ有用にすること」、「知識の創造、選択、利用を支配するプロセスの集合」あるいは「利用可能でかつ必要な情報を同定・分析し、組織の目的に適用できるように知的資産を展開するための計画や調整」など多用な定義が存在している。

これに対して、情報管理 (Information Management) という用語はより古くから使用され、広く定着している。情報が組織内で有用なものとして管理でき、あるいは管理が必要であるものとして認識されたときに、情報管理という考え方が生まれた。さらに、情報管理を実践するために情報分析や情報企画というような分野が生まれた。これと同様に、知識も組織における無形資産として考えられたときから、組織内における個々人の知識源を分析するために、各種の手法や技術を適用しようと考えられるようになった。

一方で、インターネット・イントラネットに代表される情報技術の変化に対応した企業経営のあり方への新しいアプローチとしてナレッジ・マネジメント（現在、知識管理ではなく知識経営と訳することが多いのこのような理由によるのであろう）が、多くの企業で実践されている。

例えば、Kark-Erik Svieby のサイトでは様々な角度からの多くの実践例について、顧客からの情報・知識の入手、顧客への知識の提供、既存の知識からの新しい収益の獲得、知識共有文化の創造、個々人の暗黙下の知識を補足・蓄積・拡散するための手法、知識創造プロセ

スの計量、ナレッジ・マネジメントに基づくキャリアの蓄積、暗黙下の知識を伝達するための環境の設定などに大きく分類している。

George Mason 大学のICASITでは、ナレッジ・マネジメントに関連する概念（用語）には次のものがあるとしている。ナレッジ担当責任者（Chief Knowledge Officer）、企業戦略（Corporate Strategy）、顧客資産（Customer Capital）、データマイニング（Data Mining）、IT実践の拡大（Diffusion of IT Practices）、知的資産（Intellectual Capital）、知識企業（Knowledge Enterprise）、知識生態（Knowledge Ecology）、知識転移（Knowledge Transfer）、組織学習（Organizational Learning）などである。論文やレポートでナレッジ・マネジメントについての内容を扱っていても、必ずしもナレッジ・マネジメントという用語を使用していないこともあるため、これらの用語をインターネットなどでの調査に用いると有用である。ナレッジ・マネジメントに関する個々の解説書、論文、各種事例、ニュースについては、以下のウェブサイトやウェブマガジンなどで詳しく紹介されている。

ナレッジ・マネジメントに関するウェブサイト

- ・ 日本ナレッジ・マネジメント学会 HP
(<http://www02.so-net.ne.jp/kmsj/>)
ではナレッジ・マネジメントについての定義集、関連図書あるいは学会の活動内容・予定などがわかる。最初に見るサイトとして適当である。
- ・ George Mason 大学の情報技術応用研究国際センター (ICASIT)
ではナレッジ・マネジメントについての定義、論文、図書、事例研究などを集めた Knowledge Management Central (<http://www.icasit.org/km/>) を運営している。
- ・ WWW Virtual Library on Knowledge Management

(<http://www.brint.com/km/>)

は論文、図書、白書、インタビューなどへのリンクが多く、最初に見るサイトとして適当である。

- ・ Kark-Erik Sveiby のサイト
(<http://www.sveiby.com.au>)
では定義などの他に、実際の各企業でのナレッジ・マネジメントの実践例が、多く収録されている。

ナレッジ・マネジメントに関するウェブ出版物サイト

- ・ KM World
<http://www.kmworld.com>
- ・ Knowledge Management
<http://www.kmworld.com>
- ・ CIO Knowledge Management Research Center
<http://www.cio.com/forums/knowledge/index.html>
- ・ Knowledge Management News
<http://www.kmnews.com/>
- ・ KM Magazine
<http://www.kmmag.com/>

お知らせ「論文募集」

特集「ゲノム情報学」の論文公募
Vol.10, No.4(2001年1月刊行予定)

情報知識学会誌編集委員 菅原 秀明、国沢 隆

1. 主旨

本年6月(2000年), ヒトゲノムの全解説の概要(ドラフト)ができ上がったことが発表されました。ヒトを始めとする、いろいろな生物の全ゲノム塩基配列を決定しようとするゲノムプロジェクトは、10年ぐらい前にスタートし、現在では感染症の原因となる微生物を中心として、30種以上の全ゲノム塩基配列が決定されています。こうして決定された配列データは、だれもが利用できるようにデータベース化され公開されています。ここで注目すべきは、さまざまな情報論的手法が、配列決定やデータベース化の過程は言うまでもなく、決定された配列からの遺伝子の同定、遺伝子機能予測、病気との関連、生物の進化などについての新知見の獲得に、用いられていることです。情報科学と結合した新しい生物学が誕生し発展しつつあります。この新しい分野に対して、情報知識学会の多くの会員が関心をもっていると想像されます。そこで、本学会編集委員会では、「情報知識学会誌 Vol.10, No.4」を当該テーマについての特集号として企画しました。

2. 公募論文のテーマ

下記のテーマについて公募します。

- (a) ゲノムデータベース、(b) 遺伝子同定、(c) 分子進化
- (d) タンパク質の構造と機能予測、(e) 遺伝子発現、プロテオーム解析
- (f) 上記テーマに関連するソフトウェア・ツール等

投稿規定は、知識情報学会誌の投稿規定に準拠します。論文の長さは、刷り上がりで8~16頁とします。

3. 期限および原稿送付先

投稿原稿および修正原稿の締切りは下記のとおりです。

投稿原稿締切り: 2000年10月6日(金)
修正原稿締切り: 2000年11月1日(水)

4. 原稿送付先

照会ならびに原稿送付先は、下記までお願いいたします。

〒259-1293 平塚市土屋 2946 神奈川大学理学部 後藤智範
E-mail: gotoh@info.kanagawa-u.ac.jp
TEL: 0463-59-4111 FAX: 0463-58-9684

お知らせ

情報知識学会主催 「第5回SGML／XML研修フォーラム」

情報知識学会誌10巻1号でご案内した通り、標記フォーラムを今秋主催します。2日間の内容は、行政関連、学界、ベンダー系、ユーザー系（主としてEC関連）の4区分で各4名程度、14～16名の講師を予定しています。詳細プログラムは今夏、本学会のホームページに発表します。

(<http://angelos.info.kanagawa-u.ac.jp/jsik/main.html>)

毎回好評で本学会員も多数参加されています。（参加費は半額となります。）また、当学会員が紹介された一般のかたも割引（18,000円）で参加できますので、XMLにご関心のあるかたをお誘いください。申込受付は8月中旬からの予定です。前回は定員100名に対し150名のお申込がありました。先着順ですので、参加お申込はお早めにどうぞ。ご不明の点のご遠慮無く事務局へお問い合わせください。

開催日時：2000年10月25日（水）および26日（木）の2日間 9:30～16:30

予定会場：グランドヒル市ヶ谷（東京・市ヶ谷駅より徒歩3分）

講演内容：政府・自治体等公共的場面での動向と、民間企業の適用事例および対応製品の現況を総覧

参加特典：全講演の要旨を集録した情報知識学会誌特別号（3,000円）を、当日会場受付で差し上げます。

協賛団体：多数交渉中のため後日発表

参加費：一般 28,000円、地方自治体職員・国家公務員 18,000円

情報知識学会会員 14,000円、入会申込者（当日手続き可） 20,000円

協賛団体会員・リピーター 18,000円、学生 5,000円

問い合わせ先：情報知識学会事務局 〒110-8560 東京都台東区台東1-5 凸版印刷（株）内

TEL:03-3835-5692 FAX:03-3837-0368 E-mail:LDE01013@nifty.ne.jp

なお、前回の「第4回SGML／XML研修フォーラム」詳細は、下記ホームページの”シンポジウム”欄をご覧ください。

<http://angelos.info.kanagawa-u.ac.jp/jsik/main.html>

お知らせ「論文募集」

専門用語研究部会 第13回専門用語研究シンポジウム
— 発表論文募集 —

情報知識学会理事 後藤智範

趣旨: 本年5月の本学会の総会で専門用語研究会との統合が承認されました。専門用語研究会では、設立以来毎年秋にシンポジウムを開催し、昨年度まで12回行なわれました。従来は招待講演が中心でしたが、情報知識学会の研究会として新たに発足したことであり、本年度から5月の研究報告会と同様に研究発表を募集します。どうぞ奮ってご応募ください。

開催日: 平成12年11月11日(土)、 会場: 国立情報学研究所
13:30 - 17:30 を予定しています

1. 募集分野

専門用語ひいては「用語」に関連するテーマ全般を全て含みます。具体的には下記のテーマを挙げることができます。

- (1) 専門用語理論・モデル
- (2) 専門用語の組織化、分類(シソーラス、自動分類など)
- (3) 専門用語の標準化、各国の研究開発動向、その他
- (4) 辞書編纂、専門用語の翻訳(機械翻訳など)
- (5) 用語データベース/知識ベース、電子化辞書
- (6) 専門用語関連ソフトウェア

2. 発表申し込み方法

発表論文題目、著者名(連名の場合、発表者名に○印) 連絡代表者の氏名、連絡先の住所、TEL/FAX番号、E-mailアドレスを明記の上、下記宛て、E-mailまたはFAXにてお申し込みください。

〒259-1293 平塚市土屋 2946 神奈川大学理学部 後藤智範

E-mail: gotoh@info.kanagawa-u.ac.jp、TEL:0463-59-4111、FAX: 0463-58-9684

申し込み期限: 平成12年9月15日

3. 論文執筆・発表等について

(1) 発表論文は2~8ページ(A4判)にカメラレディの形で仕上げて、(a)プリント出力したものを郵便し、同時に、(b)その元ファイルをE-mailアドレスに添付ファイルの形態でお送りいただくことになります。

発表要旨の締め切り: 平成12年10月15日

(2) 発表時間は、質疑応答を含めて30分程度を予定しています(発表者数が確定次第、正確な時間をお知らせします)。

(3) 発表者は情報知識学会の会員に限ります(当日入会も可)。

なお、標記についての最新の情報は下記のURLをご参照ください。

<http://angelos.info.kanagawa-u.ac.jp/jsik/main.html>

お知らせ

「2001 年情報学シンポジウム」論文募集
-21 世紀の情報化社会・ネットビジネスを支える情報学／情報技術-

日本学術会議を中心に、情報知識学会をはじめ多くの学術研究団体の共同主催によるシンポジウムです。今回から既発表の論文も可となりました。当日の参加費は一般：15,000 円ですが、情報知識学会会員は 10,000 円です。

ネットビジネスの広がりによって代表されるような 21 世紀の情報化社会における情報の円滑な流通と高度な利用の促進を目的とする本シンポジウムは、1984 年以降毎年開催してまいりました。日本を代表する研究者・技術者を招いて、このテーマに沿っての講演をいただき、また一般公募の論文の発表も行ないます。毎年、多くの方に参加いただき、興味深い講演、発表にホットな議論が沸き立ち、情報、知識、ネットワーク、社会という多様な側面からの知見とアイデアの交換が行なわれてきました。本年度は、上記のテーマで招待講演をいただくとともに、以下の要領で論文を募集する運びとなりました。多数の論文の投稿をお待ちしております。

日時 平成 13 年 1 月 18 日 (木)・19 日 (金) 9:30 ～ 17:00
会場 日本学術会議講堂 (東京都港区)
共同主催 (予定) 日本学術会議 (情報学研究連絡会、情報工学研究連絡会)、
情報処理学会、人工知能学会、日本医学会、日本化学会、
日本数学会、日本地理学会、日本物理学会、情報知識学会、
日本生物物理学会、日本ソフトウェア科学会
後援 (予定) 科学技術振興事業団、化学情報協会、計測自動制御学会、
国立情報学研究所、情報科学技術協会、情報通信学会、
電子情報通信学会、日本医療情報学会、日本機械学会、
日本金属学会、日本原子力学会、日本材料科学会、
日本材料学会、日本社会情報学会、日本生化学会、
日本電信電話、日本動物学会、日本農学会、日本分子生物学会、
日本分析化学会、日本薬学会

論文募集要項

トピックス：インターネットの爆発的な発展とパソコンの驚異的な低価格化は、「いつでも、どこでも、だれでも」自分のパソコンで世界中の情報資源にアクセスして利用できる環境をつくり出した。そして、通信と計算機ハードウェアの高速化、大容量化の進展はとどまるところを知らない。この流れはネットビジネスの隆盛を迎えようとする 21 世紀の情報化社会における情報技術活用のインフラストラクチャーを提供するだろう。しかし、このような環境を活用するソフトウェアはまだまだ工夫の余地が大きい。ひとつには、情報技術に対する社会のニーズが技術開発側に十分知られていないこと、もうひとつは現在の情報技術の持つ可能性が社会的に認識されていないことがある。このような問題意識の下、本シンポジウムでは、情報技術の開発側と情報技術に期待する利用者側の両方から、技術的研究成果、情報技術の応用が期待される分野に関する幅広い論文を募集する。募集する論文は以下のトピックを中心とするが、本シンポジウムの主旨に沿うより広い範囲の論文投稿も歓迎する。

(1) 21 世紀の社会において情報技術で解決したい問題

初等、中等教育における情報処理教育、情報技術の導入方法、
医療・福祉における情報技術、エンタテインメントにおける情報技術

- (2) ネットワークビジネスにおける情報技術
電子商取引、デジタルコンテンツ流通、Web マーケティング、
Web サーチエンジン、モバイル情報サービス、セキュリティ、XML など
- (3) 分野融合における情報技術
文理融合型の情報技術、電子図書館、電子博物館、農業の情報化、
異なるコミュニティ間の情報流通、言語横断型の情報技術
- (4) 21 世紀の情報化社会を支える基礎理論・技術
新しい情報モデル、知識構造、知能機構、情報検索・フィルタリング、
情報抽出・分類・体系化、情報可視化、データマイニングなど

なお、発表内容はシンポジウムの主旨の沿うものであれば既発表であっても構わない。

応募方法

ワープロで A4 用紙 8 枚以内の論文と、題目・概要・著者氏名(発表者には印をつける)・連絡先(必ず電子メールアドレスは記載のこと)・所属を明記した別紙を添えて下記に申し込むこと。なお、論文、別紙ともコピー 4 部を送付すること。また、別紙の内容をプレーンテキストの電子メールでプログラム委員長(メールアドレス: t-fukushima@cj.jp.nec.com)にも送ること。

スケジュール

- (1) 論文応募締切 平成 12 年 9 月 15 日(金) 必着
(本誌 6 月号での募集では 9 月 8 日としたが 1 週間遅らせた)
- (2) 採否通知 査読審査後、平成 12 年 10 月 17 日(火) までに発送
- (3) 最終原稿締切 平成 12 年 11 月 30 日(木)

なお、シンポジウムに関する最新情報は <http://clicks.to/sigfi2001/> に掲載する。また、シンポジウムの論文集は WWW での公開を予定している。

申込/照会先 (社) 情報処理学会 シンポジウム係

〒108-0023 東京都港区芝浦 3-16-20 芝浦前川ビル 7F

E-mail: sig@ipsj.or.jp Tel:03-5484-3535 Fax:03-5484-3534

参加費(予定) 共催学協会員:10,000 円、学生:3,000 円、一般:15,000 円

シンポジウム実行委員会: 情報処理学会 情報学基礎研究会

実行委員長: 中川 裕志(東京大学), E-mail:nakagawa@r.dl.itc.u-tokyo.ac.jp

プログラム委員長: 福島 俊一(NEC), E-mail:t-fukushima@cj.jp.nec.com

投稿規程

第1版(暫定板)1990年3月

第2版(暫定板)1995年7月

第3版(暫定板)1998年3月

情報知識学会では、SGML(Standard Generalized Markup Language)に基づく学術情報の編集・蓄積の実現を意図しています。SGMLはISO規格8879またJIS規格X4151で、文章データベース作成・保守・交換の技術として、電子出版だけでなく、マニュアルなど技術文書管理にも使われてきました。最近では、さらにネットワーク上でのコンカレント・エンジニアリングから電子商取引まで、文書情報の編集・蓄積の基盤技術として国際的に普及しつつあります。

1. 原稿の投稿資格

本学会誌に投稿原稿を投稿するには、原則として本学会会員(個人会員)に限る。ただし、投稿者が2名以上の連記の場合には、少なくともそのうち1名は本学会会員(個人会員)であることが必要です。

2. 原稿の種類

論文(原著論文)、総説、解説などがあり、原稿には投稿原稿と依頼原稿とがあります。原稿は日本語または英語で書いてください。主題は本学会の対象とすることであれば特に限定しません。

3. 査読と採否

論文の採否は編集委員会で決定し、著者に通知します。また、内容の改善を求めることもあります。論文以外の原稿については査読の結果、表現の分かりやすさなどの点から修正をお願いすることがあります。

4. 原稿の構成

表題(和文と英文)、著者名(和文表記と英文表記)、著者所属(和文表記と英文表記)、要約(和文と英文)、本文(和文または英文)、文献リスト、注、表、図など。

5. 原稿の書式

原稿のうち、テキストおよび図・表は3.5インチフロッピーディスク(1.44MB 2HDのみ)とその印字出力(2部、うち1部は査読用)で、図や表は紙(2部)でお出し下さい。

テキスト部分の文書形式は、下記の形式に限ります。いずれの形式においても文字コードはSJISに限ります。

(a) ASCII L^AT_EX のコマンドが挿入されている Plain Text

(b) Rich Text Format(MS Word、WordPerfect で使用されている書式)

(c) Plain Text format(いわゆる単純なテキストファイル形式)

使用したソフトウェアがフロッピーディスク表面にわかるように記述してください。

Vol.7, No.2 から L^AT_EX で編集・印刷しております。これに伴い、本投稿規程の変更を予定しておりますので、最新の投稿規程については下記の当学会のホームページをご覧ください。

情報知識学会 HP <http://angelos.ed6.info.kanagawa-u.ac.jp/jsik/main.html>

6. 原稿の送付

6.1 送付時に必要な書類の確認

送付時に必要なものを改めて記すので確認してください。

- ・ 標題, 著者名, 所属機関, 概要 (これらはいずれも和文と英文の両者)
- ・ 本文 (和文または英文), 参考文献, 図・表
- ・ 以上の本原稿に加えてそのコピー 1 部
- ・ 著者 1 名の連絡先 (郵便宛先, 電話番号, FAX 番号, e-mail アドレス)

6.2 原稿の送付

原稿は学会事務局ではなく、下記宛てにお送りください。封筒に「情報知識学会誌 投稿原稿 在中」と明記してください。

〒259-1293 平塚市土屋 2946

神奈川大学 理学部 情報科学科

後藤 智範

TEL: 0463-59-4111, FAX: 0463-58-9684

E-mail: gotoh@info.kanagawa-u.ac.jp

さらに、下記の内容を含むテキストファイルを e-mail で上記のアドレス (学会誌編集委員長) 宛てに宛てにお送りください。

標題 (和文、英文), 概要 (和文、英文)

著者 (1 名) の氏名, E-mail アドレス

所属機関／部課名, 所属機関の住所／電話番号／FAX 番号

7. 著作権

著作権は原則として本学会に帰属するものとします。著作権が他の機関にある場合などで、本学会に帰属することが困難な場合には協議することもあります。本学会誌に掲載された執筆内容が第三者の著作権を侵害するなどの指摘がなされ、損害を与えた場合には、執筆者がその責任を負うことになります。

専門用語研究会合併に伴う新規入会

専門用語研究会合併に伴う新規入会の方々は次のとおりです(2000年7月31日現在、敬称略、順不同)。

賛助会員(3法人)

ブラザー工業(株) 技術開発部、医学中央雑誌刊行会、(財)日本医薬情報センター附属図書館

正会員(31人)

青戸邦夫(国立情報学研究所)、崔杞鮮(KAIST(韓国科学技術院))
 藤枝真((財)口腔保険協会)、長田孝治((株)日本総合技術研究所)
 太田泰弘(文教大学湘南校舎)、石島優子(GMD 東京事務所ドイツ図書館連絡室)
 石井正彦(大阪大学)、バルバドー・ソフィー(琉球大学)
 香川靖雄(女子栄養大学)、福田滋(湘南工科大学)、吉野敬子(ウェルファイド(株))
 木本幸子(大妻女子大学)、倉島節尚(大正大学)、牧野正久(東京理科大学)
 森口稔(シャープ(株))、長尾真(京都大学)、松村多美子(椋山女学園大学)
 山本晴彦(神奈川大学)、近江晶(帝京平成大学)、貝島良太((株)日立国際ビジネス)
 寺村由比子(駿河台大学)、薄葉威士(衆議院事務局)、春山曉美、市川幸郎、高橋正美
 仲本秀四郎(IRIS 情報学研究所)、井上如、伊藤全、島地章雄、柴田武、竹森利清

情報知識学会役員(2000年7月31日現在)

会長	藤原鎮男	東京大学 名誉教授	
副会長	藤原譲	神奈川大学 教授	
	細野公男	慶応義塾大学 教授	
	根岸正光	国立情報学研究所 教授	
理事	芦沢実	日立製作所(株) 主任研究員	有川節夫 九州大学 教授
	石塚英弘	図書館情報大学 教授	岩淵幸雄 日本創造学会 理事
	菊田昌弘	(株)シナジー・インキュベート社長	
	国沢隆	東京理科大学 講師	後藤智範 神奈川大学 教授
	高橋仁一	大日本印刷(株)開発室長	小山照夫 国立情報学研究所 教授
	菅原秀明	国立遺伝学研究所 教授	田隅三生 埼玉大学 教授
	田畑孝一	図書館情報大学 副学長	田村貴代子 市川房枝記念会
	次田皓	つくばプロテオミクス研究所 所長	
	中川優	和歌山大学 教授	長瀬眞理 静岡大学 教授
	永村真	日本女子大学 教授	名和小太郎 関西大学 教授
	西脇二一	奈良大学 教授	平田周 IT 経営研究所 所長
	深見拓史	凸版印刷(株) 技術企画部長	松田芳郎 東京国際大学 教授
	安永尚志	国文学研究資料館 教授	山本昌弘 法政大学 教授
監事	宋戸駿太郎	環日本海経済研究所 研究顧問	
	月見里禮次郎	えんがわ 代表	(五十音順)

情報知識学会 賛助会員(14法人、2000年7月31日現在)

(株)シナジー・インキュベート	医歯薬出版(株)
大日本印刷(株)	科学技術振興事業団
凸版印刷(株)	駿河台大学図書館
中根速記学校	(株)トータルメディア開発研究所
(社)日本印刷技術協会	インフォコム(株)
日本ツーリスト開発(株)	(株)PFU ソフトウェアラボラトリ
(株)横河総合研究所	プロセス資材(株)
	(五十音順)

=====

◀ 編集後記 ▶

今月号から、新たに6名の編集委員を迎え、12名で編集委員会を運営することになりました。編集委員の増加に伴い、相応の投稿論文が来ることを期待する次第です。

若手の会員で当学会誌の \LaTeX 編集に協力いただける方がおりましたら大歓迎です。下記宛てまでご連絡ください。
gotoh@info.kanagawa-u.ac.jp 編集委員長 後藤 智範

=====

=====

情報知識学会誌 編集委員会

編集委員長 後藤 智範 神奈川大学 理学部 情報科学科
(E-mail: gotoh@info.kanagawa-u.ac.jp)

編集委員

藤原 譲	神奈川大学理学部	根岸 正光	国立情報学研究所
安 江虹	理科化学研究所遺伝子基盤研究部	石井 正彦	大阪大学大学院文学研究科
石川 徹也	図書館情報大学図書館情報学部	石塚 英弘	図書館情報大学図書館情報学部
国沢 隆	東京理科大学理工学部	阪口 哲男	図書館情報大学図書館情報学部
菅原 秀明	国立遺伝学研究所	中川 優	和歌山大学システム工学部
西脇 二一	奈良大学社会学部	安永 尚志	国文学研究資料館

=====

■複写をされる方に

本誌に掲載された著作物を複写したい方は、(社)日本複写権センターと包括複写許諾契約を締結されている企業の従業員以外は、著作権者から複写権等の行使の委託を受けている次の団体から許諾を受けて下さい。著作物の転載、翻訳のような複写以外の許諾は、直接本会へご連絡ください。

〒107 東京都港区赤坂 9-6-41 乃木坂ビル 学術著作権協会
TEL:03-3474-4621 FAX:03-3403-1738 E-mail: kammori@msh.biglobe.ne.jp

アメリカ合衆国における複写については、次に連絡してください。

Copyright Clearance Center, Inc. 222 Rosewood Drive, Danvers, MA. 01923, USA
TEL: 508-750-8400 FAX: 508-750-4744 URL: <http://www.copyright.com>

情報知識学会誌 Vol.10, No.2 2000年7月31日発行 編集・発行 情報知識学会
頒布価格 3000円

=====

情報知識学会 (JSIK: Japan Society of Information and Knowledge)

会長 藤原 鎮男
事務局

〒110-8560 東京都台東区台東 1-5-1 凸版印刷(株)内
TEL: 03(3835)5692 FAX: 03(3837)0368 E-mail: LDE01013@nifty.ne.jp
URL: <http://angelos.info.kanagawa-u.ac.jp/jsik/main.html>

=====

Journal of Japan Society of Information and Knowledge

~~~~~ Contents ~~~~~

Forward	1
Construction of a large scale 3D image database of human brain Kazuhito Shida, Yoshiyuki Kawazoe, Kazuya Kumekawa, Ryuta Kawashima, Masao Otsuki, Hiroshi Fukuda	2
Layers Sequences in the Materials with Perovskite-Related Structures Yuri Kotliarov, Shuichi Iwata	12
Abstracts	25
The Abstracts of 8 th Annual Conference Papers 2000	26
Reports from SIGs.....	30
Commentaries on XML	36
Book Reviews.....	42
Terminology	43
Notices	45
Information for Authors	50

~~~~~

情報知識学会誌 第10巻2号 2000年7月31日発行

編集兼発行人 情報知識学会 〒110-8560 東京都台東区台東1-5-1 凸版印刷(株)内

TEL: 03(3835)5692 FAX: 03(3837)0368 (振替: 00150-8-706543)

学術刊行物 ISSN 0917-1436