

Journal of Japan Society of Information and Knowledge

情報知識学会誌

Vol.12 No.2 (Jul. 2002)

~~~~~ 目次 ~~~~~

## 特集「情報・知識」

|            |                                                                                                                                                                                            |                                         |     |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|-----|
| 卷頭言        | .....                                                                                                                                                                                      | 藤原 鎮男                                   | 1   |
| 論文<br>(招待) | SOME ASPECTS OF THE IMPACT OF THE INTERNET ON ACCESS TO INFORMATION .....                                                                                                                  | Michael W. Hill                         | 3   |
| 論文<br>(招待) | DATA, INFORMATION AND KNOWLEDGE (DIK): Challenges in a Changing Society .....                                                                                                              | Jacques-Emile Dubois                    | 10  |
| 論文         | 適合度順検索システムの性能評価尺度 .....                                                                                                                                                                    | 相良 佳弘                                   | 22  |
| 論文         | 図書館におけるレファレンス経験の知識化 .....                                                                                                                                                                  | 岩澤 まり子、吉田 亜津美、大下 祥司、中西 陽子               | 37  |
| 論文         | Bibliometric Analysis on Japanese Advertising Activities after the End of the World War II Part I: Phenomenological features extracted from one specialized journal on advertisement ..... | Hideya Azuma、Junichi Sato、Shuichi Iwata | 49  |
| 論文         | 学術雑誌電子化支援システムとしての J-STAGE SGML 編集ツール .....                                                                                                                                                 | 中川 理、西脇 二一                              | 71  |
| 論文         | 概念構造生成のための階層関係自動抽出法に関する検討 .....                                                                                                                                                            | 森本 貴之、浅川 直輝、後藤 智範、藤原 譲                  | 80  |
| 考察         | 諸学のなかにおける情報学の位置づけ .....                                                                                                                                                                    | 平田 周                                    | 88  |
| 考察         | 出版コンテンツ配信ビジネスの提案 .....                                                                                                                                                                     | 深見 拓史                                   | 95  |
| 紹介         | 専門用語研究部会参加へのお誘い .....                                                                                                                                                                      | 細野 公男                                   | 98  |
| 紹介         | CODATA 部会関連の活動状況報告 .....                                                                                                                                                                   | 岩田 修一                                   | 99  |
| お知らせ       | 第7回 SGML/XML 研修フォーラム 他 .....                                                                                                                                                               |                                         | 100 |
| 投稿の手引き     | .....                                                                                                                                                                                      |                                         | 103 |
| 会告         | 会員氏名一覧表について .....                                                                                                                                                                          |                                         | 105 |

~~~~~



情報知識学会

特集「情報・知識」

編集担当 藤原 鎮男

本号の意図

情報知識は現代社会の活動の実体であり、活動結果の表現でもある。当学会はこの情報知識を自らの活動の対象とすることを名乗っている。そこで本号ではあらためて、この「情報知識」の現状を解析し、今後の対応指針をさぐるため、識者に意見を求め、また会員からの投稿論文のうち、関連するものを特集としてまとめた。広義に見れば、これまでに刊行された会誌はもとよりこの特集の趣旨に沿うものであるから、本特集がことさらそれらと別種なものであるはずはないが、本号は、寄稿を依頼した Hill 氏、DuBois 氏から特別、この特集のために考えて下さった結果を登載することが出来た。両氏は、情報知識の問題に長く深く関係をされた方々である。我々にとって、有識経験者の最新の意見として、熟読玩味する価値があると思われる。例えば、情報知識の関係学徒諸君、これをセミナーの輪講テキストとして利用したら有用であろう。もちろん、そこで指摘されている問題に積極的に対応されたら、会員諸公も有益な示唆を得られようことはもちろんである。

その他の論考も本特集の趣旨に即した意義をもち、会員諸氏の反応を待ちたいものばかりである。そこで取り上げられている問題は今後も本学会が継続して論じたい問題である。少なくとも、ここで提議されたことについて直ちに会員諸賢の応答があったら特集編成者として喜びこれに勝るものはない。

内容

会員の投稿論文は、いずれも情報知識の処理技術の現況の表示であると言える。それらの特徴を掲載順に述べれば、

- 1) コンテンツの検索システムの機能の評価、
- 2) 図書館ユーザーの設問のデータベース化の提案、
- 3) 電子化時代の学術雑誌編集の標準化：SGML ツールの応用、
- 4) コンテンツの概念構造自動抽出法の提案となる。

これらを学術面の研究者の活動の報告例とすれば、これに続く諸篇は、情報知識の社会関連についての会員諸氏の論考である。これらは、本学会本部役員諸氏が本特集号企画への応答として寄稿された。そもそも、本学会は、「情報知識」を狭義のソフト、ハードの関連技術学術だけを対象にせず、より広く一般社会との関連に視野をひろげ、現状の解析と今後の展開を考えようと常々主張している。ここに挙げた緒篇は、この趣旨に沿って執行部有志が抱く抱負の表明である。

平田氏は、「情報工場」を名乗り、すなわち、モノの工場が、素材を加工して製品を製造するのと同様に、素材の「知識情報コンテンツ」が組織化された現代においては、この素材を製品化する「工場」機能の構築が、現代の情報・知識社会の課題であると主張し、実際活動をすでに始めておられる。寄稿の論考はその経験にたつての意見である。

深見氏も同様の立場にたち、とくに、コンテンツの蓄積と流通の整備が進んだ現代において関係業界人の最大の課題の一つは、その「配信路」が不整備であるとする。そして、この視点によるビジネスの創造が切望されると主張される。著者自身もすでに、この線に沿った実際活動を展開中である。その意味で、上記二篇は本学会活動の広がりを示す例と言えよう。続く緒篇は、目下、学会が努力中の部会活動の報告である。

以上は情報知識社会、あえて言えば情報知

識学会の現在の活動の表現である。これに加えることとして、ここで、必須なのは情報知識の本質論であろう。二篇の招待論文は、まさしくそれを示してくれている。それはまた、本特集の編集者が著者である両氏に説明した特集の意図に対して、正確に応答して下さったものであった。感謝であり、我が意を得た思いである。ここで両氏の履歴を簡単に照会すると以下ようになる。

周知のように、情報知識社会の変革は、計算機の発達による科学技術とくに医学、化学、生物学などの学術情報の機械処理に始まった。英国図書館はその先導者の一つであり、Hill氏はオックスフォード大学の化学修士の学位を持つが、早くより図書館改革の渦中であって実際に学術雑誌流通の改革の中心の役割を果たした人物である。英国のブリテイッシュライブラリの一部である Science Reference Library(科学図書館)の館長職をつとめ、また国際情報ドキュメンテーション連盟(通称 FID)の理事、会長をされた。小生は同氏と同じ時期に理事、副会長として10年親交を得た。Hill氏は、本稿で、学術情報は情報知識社会が現代のように拡大拡散しても、なお、その中枢としての地位はゆるがぬであろうと断言し、なお、現在よび将来において情報コンテンツ流通の課題は、旧来確立された「論文(情報)の質の評価法」の代替えをいかにするか、コストの分担をどうするかであるとし、さらに本源の問

題として、創造的研究支援のための情報知識のアクセスのあり方について、さすが、英国でその中枢の責任に当たった方の意見であると頷かせる卓見を披露された。

DuBois 教授は 1974 年国際純正応用化学連合(通称 IUPAC)に化学情報部会ができたとき、その委員長となり、米国ケミカルアブストラクツ(通称 CAS)の Dr. Tate とともに化学情報の機械化、画像化の実現に努力した。自身、パリ第7大学教授として、化学情報処理システム、DARK、をつくり、またフランスの軍産官学を網羅した共同研究グループを組織してフランス全体の情報センターを構築した。いささか私事に関わり恐縮であるが、同氏は、当時我々が構築に鋭意努力中であつた「東大情報検索システム、通称 TSIR」の内容を非常に高く評価し、先導者とすら遇してくれた。筆者は同氏が IUPAC の化学情報委員会を創設された当時から委員として協力し、その縁で、今回の特集にも特別のご好意を得た。ちなみに本稿は今秋同氏が米国の 20 世紀の情報学の顕著な研究者として招待されている講演の予稿に即した内容のものと聞いている。本稿では、同氏は情報、知識を文字、画像、ファクトの三成分から成るという視点で、それぞれの本質を把握して、かつ、その総合で情報知識の問題にあたるべきことを述べておられる。読者の熟読玩味を期待したい。

SOME ASPECTS OF THE IMPACT OF THE INTERNET ON ACCESS TO INFORMATION

Michael W. Hill

Formerly Director, Science Reference Library, the British Library

We live in interesting times. There is no peace. Before we have adjusted to one new technology another is upon us and we have to start the adjustment process all over again. A period of calm in which to resolve how our needs can best be met is denied us. We are robbed of sufficient time to experiment thoroughly and systematically by improvements to each new technology of such frequency and magnitude that the equipment or system is obsolete before the research is finished.

A slight exaggeration perhaps, but not much, particularly when the modern information and communication technologies (ICTs) are being considered. Nevertheless, it is necessary to try to stand back for a moment from the forefront of progress and try to judge where the latest ICTs are leading us, what the consequences are likely to be and what changes we must make to how we work in order to gain maximum benefit and minimum disbenefit.

If we look back at the history of ICTs, we find that previously, after the excitement of each new technology, there has come about a rebalancing of all the then existing ones. Even today recording on stone is still used but only for static monuments. Printing did not eliminate manuscript, nor did the telephone, though first typing and now e-mail are greatly reducing the number of manuscript letters. We must, therefore, anticipate some redistribution of tasks for which each of the following is used:-

engraving	voice
manuscript writing/typewriting	smoke signals
printing	flag signals
microfiche/microfilm	telegraph/telegram/telex
television	radio
tape recording	fixed telephone/fax
CDs	mobile phones/text messaging
Internet/web sites	e-mail and Intranet

Of course, a technology can serve more than one purpose. This is particularly true of the Internet, which is only a system for interconnecting a vast number of computers together on a world-wide scale. Thus it facilitates correspondence to be sent simultaneously to many recipients, it enables computer to computer discussion groups to be set up, and it enables remote stores of information, text and pictures to be interrogated. For this it is necessary, of course, that all the communicating equipment and systems are compatible. Therein may lie one of the major problems of the future. Text stored using today's equipment/systems may prove inaccessible to the equipment/systems of the future.

It is an oft repeated platitude that the Internet enables all the peoples of the world to

communicate with each other and to access all the information in the world. And, like all platitudes, it is only partially true. For those who can afford a computer and can learn the skills, it brings huge benefits but one does not use it to the exclusion of other ICTs. For example a telephone allows the subtle shades of meaning conveyed by the nuances of the human voice to come into play. E-mail is a very impersonal method of communicating compared with a letter, especially a manuscript one.

Nevertheless, the e-mail facility that the Internet allows is, on the whole, beneficial as far as the dissemination and growth of information and knowledge is concerned. It is much faster than postal services and unlike the telephone it does not require the recipient to be present at the time of sending. Answerphones, though useful, are not an adequate alternative. It has the huge advantage that the same message, or request for information, can be sent simultaneously to a large number of people in the expectation that, at least one will respond with the answer wanted. Experts may, however, find this a nuisance if they get too many requests, especially from schoolchildren undertaking "projects".

Unfortunately, e-mail has also raised a new barrier. Many newsletters that were previously printed and distributed free are now distributed only as attachments to e-mails. As a result those who do not have access to e-mail facilities are deprived of an information source that once they had. In industrial countries libraries provide access for the poor to e-mail and the Internet generally. In general, this is not so in the poorer developing countries. Fortunately, even in the poorest countries, there are some organizations which have acquired or been donated means to access the Internet, though without back-up electrical generators there may be problems from time to time of power supply. Among such organizations are hospitals and universities. For these arrangements are being made by UN and some professional bodies in industrial countries to provide free on-line access to publications. One huge advantage of the Internet is that the cost of providing access to one more customer is trivial.

Quality Problems

When we turn to the question of the value and reliability of information sources accessed through the Internet, interesting questions arise, especially if we try to compare these aspects with access to equivalent sources by more traditional means. For this it can be helpful to divide the use of the Internet into its three basic modes:

- (a) person to person, whether one to one or one to many;
- (b) person to known website;
- and
- (c) person to Search Engine.

In the first case the situation is no different from when one is seeking information by letter or telephone. This applies whether the information sought or gained via e-mail, Discussion Groups or Bulletin Boards. In each case one is getting information from individuals and one simply uses one's own judgement of the knowledge and reliability of those individuals.

In the second case one has to rely on the website but here too one's experience comes into play. If the website is that of a distinguished professional body one would accept what one finds just as one would information it had put out in print. So one would do also with a newspaper's

website. One judges what one finds on a website by the reputation of its owner

High calibre scientific journals submit new research papers to the peer review process, but even this is not infallible. Errors are revealed by subsequent research and the correct results then published. Searching for a subsequent paper in printed journals is greatly assisted by Citation Indexes. One hopes these will continue at least until searching electronic publications by key words in context or by citations is proven to be at least as efficient.

It is when one uses a Search Engine that problems may arise. These Engines do not validate information. They do not institute any process of peer review. They simply search for terms in all the stored text in computers to which they have access and many people and organizations have made a vast mass of information, ideas and beliefs available to those who search the Net. One could argue that a vast mass has similarly been made available in books and journal articles but in these cases there has usually been a measure of control. Publishers have always been wary and even if peer review has not been instituted their editors have been watchful. Self publishing is expensive and thus the amount of nonsense published in conventional format is limited. The Internet, on the other hand, allows publishing at very small cost and as a result not only sound material is published but also, sadly, a great deal of material of dubious validity. A search for information rarely reveals who is the source of the information accessed. Those who search need to be wary and when there is need try to establish some way of checking the reliability of the answer. But one must not exaggerate the problem. For many searches the result is quite acceptable, simply using one's common sense to test the likelihood of the result being right.

Thus when today I searched via the Search Engine "Google" to find the date when the UK Freedom of Information Act was passed, and found the answer December 2000, my knowledge of the topic led me to believe this to be probably correct or I could then cross-check with a list of Parliamentary Proceedings for that date. The main problem was the vast amount of related but irrelevant information that the search revealed. Fortunately what I wanted appeared with very little scrolling.

One website which provided an extreme example of unreliable information was the Friends United Site. It aims to enable former school and college friends who have many years ago lost contact to renew contact. Anyone wishing to contact former friends simply posts his personal details with a short message. Unfortunately, one person used the facility to post a libellous message about a schoolmaster who had taught this person and his friends. The message has now been deleted, but how many more deliberately erroneous messages have been placed on websites? As everyone knows, there are many people who will not accept the conclusions of scientific research and insist that contrary statements must be taught in schools. Do they not also disseminate their beliefs via the Internet? Indeed, it is well known that there is some information accessible via the Net which is accidentally erroneous, some which is the result of erroneous belief, and some which is deliberately erroneous but, making searches generally worthwhile, much, much more which is correct.

Making new research accessible

Already the Internet is changing the way research is published and is changing the way the researcher finds records of research conducted previously. It is, therefore, important that re-

search workers generally clarify exactly what sort of information they need, in what form and how they can get it most efficiently so that they can provide clear input to publishers, professional societies, information professionals and others who will be involved in the Internet-based publishing chain.

As is only too well known, the traditional system of printed journals has become less and less efficient and more and more cumbersome and expensive as the amount of research has grown. It has also become messy. In the second half of the 20th century, the normal pattern of publishing has been first a Note to Nature followed in due course by the full paper in a journal. Later still an abstract appeared in one or more abstracting journals and in on-line services. If one did not have the relevant issue of the journal nearby, one had to get a photocopy. But, not all new papers appeared in journals. Many were published as conference proceedings, copies of which could be very difficult to track down. Some appeared in Report Series. For social scientists official publications and semi-official ones were important and so too were a wide range of ephemera. Tracking information was getting more and more difficult.

Apart from the delays between submitting a paper and its appearance in print, and the increasing costs of journals, there was the added irritation that the journals carried a wide range of articles, many of which were irrelevant to the researcher's interests. Speed of availability, ready access and ability to identify immediately relevant papers at a low cost became increasingly important.

Berners-Lee conceived the World-Wide-Web as a means of achieving immediate publication, overcoming the lengthy delays that had become normal in the print-on-paper journals but, I believe, he did expect the electronically published articles to be subsequently published in the conventional, peer-reviewed journals. Publishing on the WWW enabled those currently researching in a field in widely scattered laboratories to keep right up-to-date. Any strange results could be immediately queried.

Immediate access to current research by colleagues elsewhere is not the only information need researchers have. When starting a project one first wants to find out what is already known and what work has been done so far. During the research one will need all sorts of data, for example chemical, physical and biological properties of materials. One needs to read around one's subject in order to get ideas, which often arise from other disciplines. One may need news of new apparatus, new inventions and new materials which might be useful. In some disciplines, research published many years ago is still valuable so some form of archival store is vital.

It is important, when considering how information is to be accessed, to keep in mind that a great deal is acquired during informal discussions with fellow research workers, not just those in one's own field of work but also those on other research and in other disciplines. Commonly this occurs over coffee, at lunch or dinner or over a drink in the bar even more than in the laboratory. While there is no doubt that e-mail and Internet discussion groups are supplementing this by encouraging informal contact with researchers in other laboratories and may partially replace some formal conferences too, I doubt if face to face contacts in a sociable setting will ever be entirely replaced by the screen. Such modes of gaining information and exchanging ideas are not merely a result of the frustrations that paper-based sources of information can arouse.

The Internet has already led to research papers (and much else such as newspapers) being

published both in paper copy and in electronic version. Already some publications appear in electronic version only. If the paper is submitted electronically the only delays should be any editing and peer review and even those should be much quicker than before. Costs exist and have to be met. Will advertising on the site provide much revenue? One doubts it. Two alternatives exist. First the obvious one that users have to pay an access fee and second that the author pays a fee. This is the system that the publisher Biomed Central is adopting.

A crucial aspect of using articles on a publisher's website is that searching using keywords shall reveal only wanted papers. It must be very doubtful if the labour intensive abstracting services will continue very much longer, even as on-line services, when direct searching becomes the norm. One does not want to have to sift through all the unwanted articles as one did with the printed journal. Also there should be no undue difficulty in the way of reading the paper and getting a print out. Paying a fee for a copy must be simple for a once-only user.

Searching using keywords in one's own discipline should rarely create a problem. Difficulty may arise when searching in a different discipline in which one is not familiar with all the technical terms. There will be a need for good glossaries. These should be available on the Net so that one can find the term which expresses a concept, a task difficult with printed glossaries in which entries are all under the term.

Archiving

In general I am optimistic that as problems arise from increasing dependence on the Internet and access to sites researchers will find answers. One project to tackle some of the questions is already underway known as the Budapest Open Access Initiative (BOAI). One approach it is examining is the creation of new, non-copyright journals. Another is self-archiving.

However, there is one major anxiety in my mind which has arisen already in the case of on-line abstracting services. It is that only a limited retrospective data-base has been stored. When one attempted to search back in on-line data-bases one often found that only about 20 years retrospective entries were stored. For earlier entries one had to use the printed abstracting journals. That was not very serious because the printed issues existed, stored in big national and academic libraries. However, if research papers in future exist only in electronic form, how shall we find back copies if only those of recent date are kept on the website?

Whatever system of publishing new research emerges, it is vital that back research should be accessible. It may be that the national libraries or some other organizations of guaranteed permanence should be required to accept the role of storing archivally records of research. Other than printing paper copies and distributing them to several such centres, it is difficult to see at this stage what reliable system can be set up. One hopes that the BOAI will see a way forward. Until one is found, the learned and professional societies must be urged to maintain their printed journals as a service to their members, archival copies being deposited in national and academic libraries.

抄訳

インターネットの情報アクセスへのインパクトの諸問題

マイケル W. ヒル (元大英図書館、科学資料図書館長)

新技術の開発が相次ぎ、それへの対応に追われる現代は、面白い時代であるが当事者にとっては心の休まる時がない時代である。その当事者が直面している問題をあげてみよう。

問題は ICT (Information Communication Technology 情報通信技術) にとくに顕著である。本稿は、この ICT についての変化の状況を先ず見ることにする。

まず ICT の変化の歴史 (本文参照)。

現代 ICT の一つの特性は、それが一目的対応に限定されないことである。このことはとくに全世界の膨大なコンピューターが接続するシステム、すなわちインターネットで顕著である。これは実は、すべての通信機器、ないしシステムが互換性をもつことを要請しているのであり、これが将来の根本問題となる可能性を孕んでいる。すなわち、今日のシステムないし装置に入力されたコンテンツが、明日は取り出せなくなる可能性があるのである。

インターネットは個人と全世界との会話を可能にした。その便利さは巨大だが、全能ではない。例えば、電話は音声の微妙な意味付けを可能にする。同様に、文字の文章や手書きの手紙も E-mail では不可能な人間の息づかいを伝える。もちろん、インターネットが量的に拡大化するばかりの情報の流通に果たす利便はいうまでもない。とくに一つの通信を多数人に同時に発信するとか、逆に受け得るのは大変な便利である。しかし、それにも限度があり、質問が殺到し、それへの応答が要求されるような場合には用をなさない。

インターネットのもう一つの不便は電子機器の装置の普及の問題である。開発途上国などにおける不便さは大きな問題として残っている。とくに電気のない地域の問題は深刻で、病院や大学などでこれは大きな問題である。国連その他の援助があることは記録さるべきである。とくに、それで出版物のオンラインアクセスが可能になる意義は大であり、それに比べればアクセスコストが小さいことである。

情報の質の問題：

インターネットによる情報取得における情報源の価値と信頼性は大問題である。これを (a) 個人対個人、(b) 既成 Web サイト利用、(c) サーチエンジン利用の 3 つのモードに分けて考えて見よう。(a) は E-mail を使うとしても、評価は検索者自身でしなければいけないから旧来と変わらない。(b) は Website の能力に結果を依存する。だから、Website の能力、その評価にたよることになる。問題は (c) である。ここでは評価なしの情報が大量に出力される。もちろん旧来の書物や学術雑誌から得られる内容も同様に大量であるけれども、後者の場合は質についての査読を経ている点が根本的に違う。サーチエンジンの場合は情報源すら明らかでない場合すらある。ただし、この点を誇張すべきではない。要は注意深く検索者が評価すればよい。それには、別の項目名によってクロスチェックするのがよいであろう。

新研究情報の検索：

インターネットが研究者の発表手段や既往の研究情報の取得法を変えたことは確かである。World-Wide-Web は研究結果の即時周知 (immediate publication) を可能にした。このことの本当の価値は、普通言われるような、遅い旧来の学術雑誌の刊行の遅れを無くしたことではない。そうではなく、電子出版の結果、研究室の研究者が現在やっている研究対象について、即時的に、

世界に広く分布する関連研究者の研究現況を、指標ジャーナルで把握出来るようにしたことである。例えば、そこで不審な妙なことがあると気づいた研究者は、直ちに当人に照会する利便を得るのである。

ところで、研究者にとって必要なのは、他人の研究の様子を即時的に知ることだけではない。新たにプロジェクトを始めるときは、これまでに何が分かり、どんな研究がなされたかについての知識が必要なのである。その時役に立つのは、自分の専門外の分野での新装置や新手段、関連する物理、化学、生物学の知見、そのアイデアを知ることであり、しばしば、その古い記録が有用なことが起こるのである。それ故、保存記録の価値は無くならない。

研究者にとって役に立つ情報の大部分は、同僚、それも同じ実験室の人だけでなく、違う分野の人との特に非公式な場での会話で得られるのである。それも、同僚とのコーヒーや、ランチやディナーや、ラウンジの雑談でも起こるのであり、実験室内だけの生活からは生まれえないし、E-mail 情報からも生まれえないものである。face to face contact の意義が失われることはないであろう。学術誌の出版の遅れで情報の獲得とアイデアの交換が出来ないとよく言われ、電子出版がその欠を埋めると言われたが、それは違う。

電子出版は時代の趨勢である。その場合のコストを誰が負担するかは今後の問題であろう。広告に依存する策は疑問。可能な対応は二つである。利用者負担か著者負担。後者は出版社の Biomed Central がすでに採用した。利用者負担は現実無理がある。事項検索で得られる情報は膨大で、一々それに費用を支払うことを単純な検索者に求めるのは無理がある。

専門分野のキーワード検索から新問題 (研究対象) を生むことは考えられない。さらにまた、自分の専門分野でない分野で慣れないキーワードで検索しても好結果をすぐ生むとは思えない。この場合は、グロッサリー (総合検索語集) 支援が欲しい。それには、事項の概念を指示してくれるグロッサリーをネットが供給してくれることが望まれる。このグロッサリーは固定ターム (事項) で構築された印刷体では作れない。ネットの機能に期待せざるを得ない。

資料保存

筆者は、研究者がインターネット依存に走るために起こるとされる問題には楽観的である。研究者自身が解決するだろうと思う。ブタペスト オープン アクセス (Budapest Open Access Initiative, BOAI) は、指摘されている問題を解決する試みとしてすでに始まった、新しい、コピーライト無縁ジャーナル収集の例である。これに代わるもう一つは、自分でやる方式 (self-archiving) である。

筆者はオンラインアブストラクトについて大きい危惧を持つものである。現在のデータベースはわずか 20 年分しか遡及出来ない。それ以前の分は、資料が必要となれば、国立中央図書館や大学図書館にゆけば事が足りる。しかしながら、今後、資料保存が website だけになったときは、遡及検索はどうなるか。心配せざるを得ない。

新研究の出版がどのようになろうとも、過去の研究の知見を知り得るようにしておくことは絶対必要事である。研究記録の永久保存を使命とする国、ないし、それを保証する図書館の建設が必要である。印刷体以外のコピーをこの主の図書館に配布するシステムが出来ようとは思えない。BOAI がその責めにつくことは望まれるが、それが出来るまでは、当面、諸学会が印刷体を維持し、これをしかるべき保存図書館に寄付し、保存をはかることを熱望してやまない。

DATA, INFORMATION AND KNOWLEDGE (DIK): Challenges in a Changing Society

Jacques-Emile Dubois

President of CODATA France*

Professor Emeritus of University of Paris 7

ITODYS: 1 rue Guy de la Brosse, 75001 Paris

Dubois@paris7.jussieu.fr

In this issue of your Journal of your Japanese Society on Information and Knowledge, I wish to highlight the great changes that occurred in the last fifty years. As the history of Science and Technology did not begin during that period, I want to underline the characteristics of the classical approaches to Science, resulting from a long history of pragmatic and intellectual contributions.

The 19th century introduced numerous changes in our knowledge whose effects converged in the present Information Revolution which is, in reality, a Knowledge Revolution.

After a first section on ST aspects, I think it proper to elaborate on the structural changes in Information Circulation, first within the ST community and then in the world at large, via the present and future communication frameworks.

Both the ST world and the real world have changed dramatically. Still greater evolution is to be expected and, in this evolving situation, it is interesting to examine the future role of CODATA, its goals in ST proper and at regional and national levels.

Science and Technology

In the last fifty years, our vision and understanding of the physical world has changed drastically. We have more and better data nowadays, but the problems to be solved have become more complex and difficult. It is worth while scrutinizing the factors responsible for this evolution of science and for the present situation where science is ever more embedded in our societal changes. Two important centuries of science (18th and 19th), during which scientific disciplines were identified, led to what is called the Galilean and Deterministic Revolution. The first half of the 20th Century saw the introduction of Quantum Mechanics concepts based on the uncertainty principle at the microscopic level and on the Theory of Relativity for modelizing at a “universal” level. The second part of that century was also extremely important for various reasons.

In the 1960s CODATA was created to deal with problems emerging in the data fields, such as data quality, coherence of data files, ... It was urgent to develop the search for standards, canonical semantics and coding grammars to facilitate communication in a world in which a great need for homogeneous data acquisition and specific data was felt. Most data originated from scientists and engineers. That primary information was screened by a complex but efficient system of scientific Unions, with or without some Abstracting offices and publishers. Their

* Post President of CODATA

management, based on peer systems, was efficient but slow. Worldwide, one had to deal with quantities of data whose volume grew by leaps and bounds. Increasing quantities of data, a greater variety of good quality data, justified great changes in “data-information-knowledge” systems (DIK Systems).

In fact in that second part of the 20th century, the computer impact, both in data handling and in data instrumental production completely altered data handling processes.

To deal with such an evolution, CODATA introduced selected activities into its data policies, oriented to computer software and to the graphic and visualization aspects, approached with interdisciplinary insights. Other discoveries also induced important changes and led to new sources of data. Thus the optical (laser), mechanical and instrumental revolution gave rise to new accurate measurements in older fields with tremendous advances in the accuracy of finer measurements (using reliable probes for measurements in various fields, e.g. medicine, materials, geographic mapping and astronomy (satellite)).

During those fifty years, progress can be better understood if one considers that new information handling makes it possible to use new tools and led to interdisciplinary and multinational projects (such as GIS tools for precise geophysical surveys). Scientists began to deal with difficult problems or so-called “real world” problems and scaling factors.

In the CODATA Conference in Baveno in 2000, I found it useful to propose a logarithmic scale SLS (Fig. 1) that can, with the proper units, be used to locate measurement ranges.

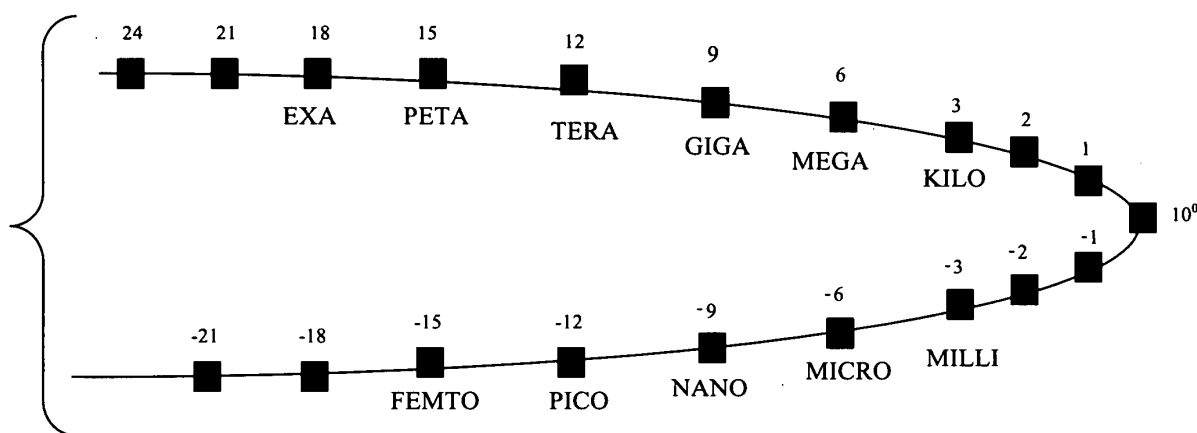


Figure 1:

Any improvement concerning the lower limits of a measurement in a given field leads to the virtual production of many more data. For instance, measurements in physical chemistry within $10 \exp -12$ seconds led to the production of kinetic values of the order of $10 \exp +12 \text{ sec}^{-1}$ kinetic constants. One can easily locate both the reaction measurement size with its scale used for different techniques, as well as the reactivity constants within the positive upper branch s. Comparison of scales is thus made more visible. Any specific scale or domain can be embedded in the GLS if one decides on proper units.

For instance, the Hubble Space Telescope archives produce 1 Giga/day, and the corresponding database is in the terabytes; the “Hypeccus database” (satellites) is expressed in Petabytes. In fact most sciences have to deal with different scales of data delivered by different access

technologies. The world-wide domain covered can be mapped for comparison on this General Scaling Logarithmic Support, useful for representing data location from the “infinitely small” to the “large worlds”, formalized either by the Galilean or Quantum theories and, more recently, with complexity methodologies.

The logical approach of more natural systems (ecological, real world, oceanography) is based on extensions of the original physical sciences developed with organized simplicity assumptions, but the progress underway stems from many contributions, some of them not strictly scientific. It has been proposed to deal with them within a formal general science of information (Tosio Kitagawa’s view), but in fact, they are developing through a gradual blending of concepts and tools from various horizons. This corresponds to the emergence of certain types of logic, e.g. fuzzy logic, associating rational ST infrastructure or abstract systems and other aspects of human activities and human culture.

Gradually, our comprehension of the world is more ambitious, and the Information Society corresponds, in fact, to a society paradigm implying that our goals and ambitions are broadening and computerized. The “Infinitely Small”, dealing with invisible entities, is based on the molecular paradigm and deals with biogenetic challenges. The “Infinitely Large” relates to the explanation of our Universe (origin and human action). Information Concepts rely on Knowledge Systemic Issues, human cognitive aspects, DIK relations and Communication and are instrumental in both “small” and “large” spaces (Fig. 2)

Thus, from simple models our sciences gradually cope with real situations presenting multiple choices. In other words, the last fifty years correspond to another scientific revolution accepting

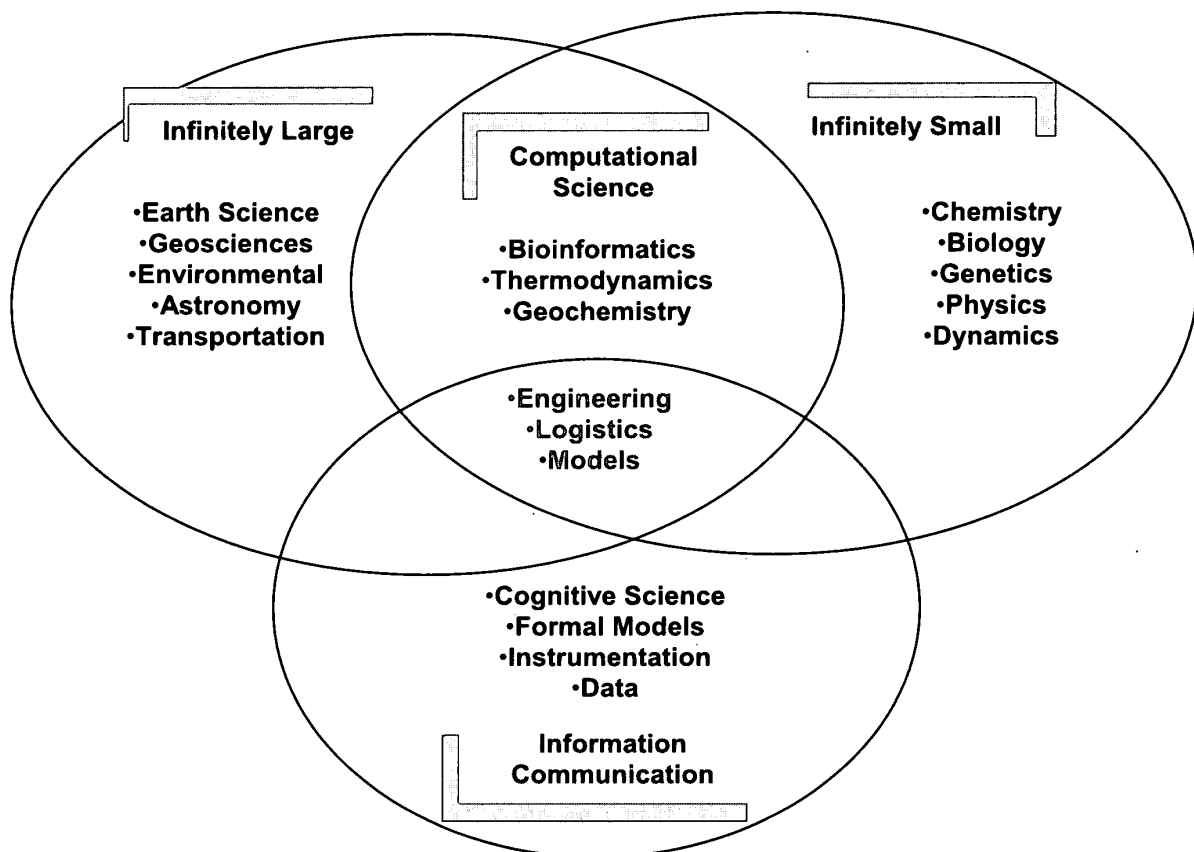


Figure 2:

the complexity paradigm as a normal basis, the “irreversible processes” in the forefront of science even at the macroscopic level, and the search for “true” representation of the real world with adequate new mathematical tools.

For example, the large amounts of data in geophysics can be approached with modeling tools, involving the most recent mathematical tools developed around the principles of fractal structures, deterministic chaos, fuzzy logic, wavelet theory and strange attractors (Fig. 3). Even the Fourier transform, revisited in 1965, provides powerful methods for investigating linear dynamics. It is impressive to witness the rapid expansion of the latest mathematical and statistical tools in all ST communities (Fig. 3).

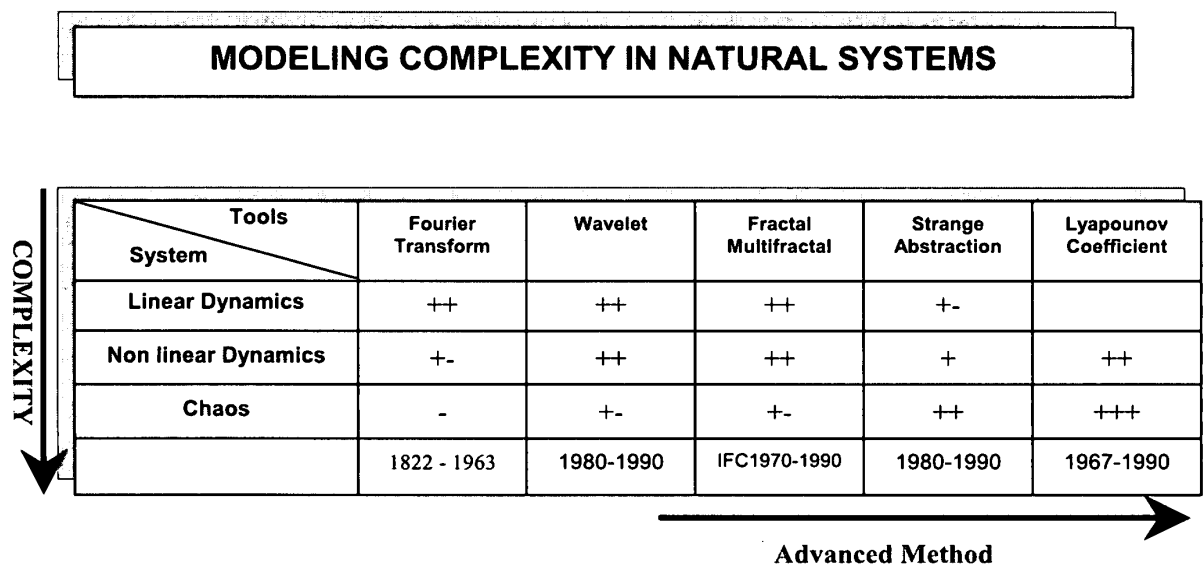


Figure 3:

In the usual development of ST, our representations have been simplified by the limitations of our knowledge concepts and theories. The Information Revolution opened a new period where progress is such that we can now face a certain complexity without degrading its problems to their simple limits. Essential features of the IR affect human cognitive progress, renovate communication and information transfer. Thus, it has deep bearing on, and questions for, our learning methodologies.

DIKW and Communication

Through this partial presentation of the “complexity” of our present day natural, societal and scientific problems, we can appreciate the increasing ambition of current issues and projects. The interval between new theories and concepts and their implementation, often by software programs, is far shorter than ever before. The transfer of theories from one field to another is important. The success of imbricated tools is partly linked to the pressure of the new flux of data. At present, in one day, the volume of data in a single field collected worldwide is much greater than that collected over decades in the past. To control this influx of data, new metadata are generated to master the handling of DIK. Adequate ontologies must be adopted

to organize data into information, both for transfer and communication as well as to facilitate the emergence of expected knowledge and innovating knowledge.

In reality, the interconnection within the DIK cycle depends on the quality of the three aspects: D, I and K and their respective tools. A population of molecular data collected for correlation is usually organized at best for a special correlation of the K design process. It is difficult to use older sets of data not tailored for a given model.

Similar constraints appear between D and I, and this is becoming crucial at the Server level, via Internet, in areas requiring rapid treatment.

A succession of parallel and integrated progress, considered as evolutionary steps, sometimes born of a real discovery, e.g. the laser, or from the application of an elucidating program, e.g. neural network, are examples of the components of the complexity evolution. They form part of our vision of what experts call the Information Revolution.

Since it touches upon so many aspects of progress, one comprehends how it readily affected the basic architecture of our science organization inherited from the past century. Present primary information is not injected systematically into the slow peer system but transferred, if well shaped, via internet long before it becomes secondary information.

In parallel, a new sector called TIC (Technical Information and Communication) aims at integrating the teaching procedure (E teaching) as a whole with the new multimedia programs and the professional software available for a particular research or service.

Data within the DIK Cycle

Over the years, the basic definition of data has changed. We recall this basic definition in Fig. 4.

In fact, nowadays nobody considers them as isolated items. They are dealt with within the DIK relationships, themselves embedded in a larger representation, DIKW, including the

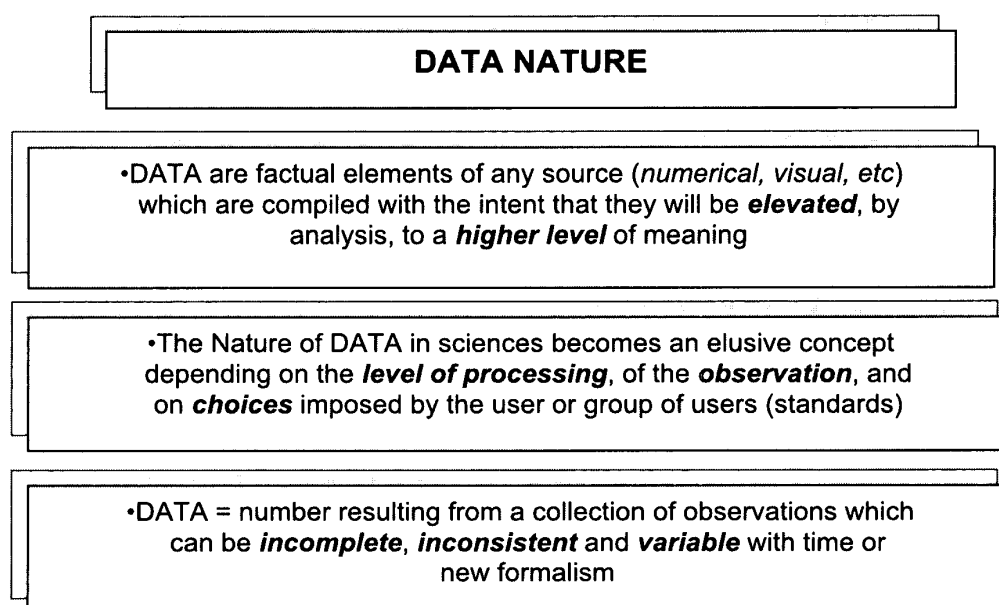


Figure 4:

discussion of complex results at the wisdom level (W) within a Knowledge Management System (Fig. 5).

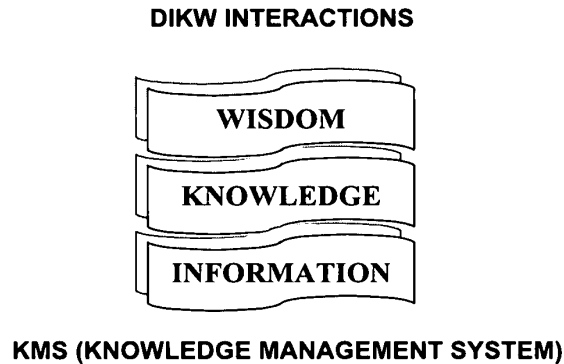


Figure 5:

Such a KMS maintains its performance through its systemic computerized handling of all types of DIK items in different presentations allowing for some visualization steps.

Hard and Soft Data

The new situation created by technological and conceptual advances evolves drastically into new fields of soft and fuzzy data. In complex cases, human and sociological factors are important (sustainable development in environmental studies). Our interpretation of systems is complicated even in the scientific world when one goes from molecular chemistry to pharmacological interactions and even medical applications. Complex elucidation systems are required (fuzzy logic, neural networks) to handle hard and soft data within in model.

In many “real world” applications, the intricacy of more precise data and knowledge operators create complex recursive problems that often often lead to a set of solutions rather than to a single one.

The different facets of a project for elaborating a solution or an object in a given time, often requires team work and the ability to cross information from different sources and of different importance. Often a team at the cross section of several fields has to use diverse “language knowledge”, adapted to mathematical tools in an optimization process. Problems of taxonomy and ontology, special vocabularies, arise in these conditions, as well as new standards involving fast decisions.

Standardization of basic data and metadata for their structuring, are essential for delocalized work in virtual laboratories or multi-expert projects. No DIK system works without parallel progress in communication tools, allowing for interoperability of distant and/or heterogeneous databases. The necessary transfer of data and protocols are hampered by the non total operability between sites. Often limitations occur in models through such handicaps. These aspects of DIK handling imply numerous standardizations and agreements affecting the chain: “raw data - deconvoluted data? metaorganization? design steps”.

International, local and cross discipline data are needed by information experts at the basic and intermediate levels. One should not be surprised that this cycle concerns the education of

the new generation. It already alters our own working habits.

It is interesting to realize that our ST society is now growing closer than ever to Real World problems. Our society as a whole is adopting an integrated view and a similar approach for societal endeavors. At different levels, the basic analysis of situations and the search for solutions are examined with DIK tools and handled with a Knowledge Management System. CODATA has been an active pioneer in this field to improve the ability to conduct simultaneous actions in many fields. Synchronization in team action is necessary to obtain a good transfer of all DIK components at DIKW levels. Fig. gives a bird's eye view of the complexity and integration needs of many disciplines and fields.

Within societal complexity, some specificity of human interactions must be taken into consideration. CODATA with ICSU was thus involved in the adaptation of Intellectual Property Rights with the IR evolution.

Moreover, advances have also involved us with security, legal and ethical issues, e.g. generic privacy. Thus, scientists are now concerned with the consequences of the impact of their discoveries on certain aspects of life and society, e.g. environmental pollution. CODATA, with ICSU, has been active on these fronts and, during its Baveno Conference, organized a successful workshop on Intellectual Property Rights.

Codata Orientations

As an international organization, CODATA has been consistently active in many of the new fields. It has maintained a good balance between classical and new ST activities and Data/Information Processing Developments. In fact, it has helped in constituting reference systems in chemistry, materials and biology as well as in collecting information.

Bridging fields, CODATA has backed interdisciplinary action by dint of various Workshops and Task Groups, e.g. pharmacovigilance and bioinformatics, physical chemistry and biology, industrial design and materials, computer design methodology. It was also successful in bringing together interesting experts in data from different horizons. The Task Group on Data Sources in Asian-Oceanic Countries has gradually built a broad-based and efficient multi-scientific community engaged in collecting valuable information and establishing contacts throughout that vast region.

National Activities

In reality, the scope of activities has been so broad that National Committees have been extremely successful when they specialize and become leaders or co-leaders in certain fields, careful nonetheless to maintain contacts with the whole spectrum of activities, usually through Task Groups and Workshops held between the Biennial Conferences.

The American Committee centered its activities on IPR questions and Environmental Issues. The French Committee has dynamic bilateral activities with other National Committees, primarily on:

- Mathematical Methods (fuzziness, chaos, molecular graphs) in the earth sciences and on

the borderlines of chemo-informatics (France, Russia).

- Interoperability and Transfer Problems in Networks (France, U.S.A.)
- Visualization and Museum Collections (France, Germany, U.S.A.)

Active participation also exists between France and Japan for the two last named objectives.

In all the above, success was achieved mainly because we were fortunate in having excellent key academics as well as industry experts in these areas.

Prospective

Promising fields of activity require attention. National Committees should identify their work with these goals. I will only mention one example here, that of the Gas Hydrates Project (CODATA/ICSU). Thanks to a worldwide investigation involving many disciplines, this should lead to information for a better understanding of gas hydrate stability so as to prevent accidents occurring during drilling, transportation or even during the exploration phase.

Today's Trends

In Advanced Computing, parallel visualization using Grid Technology is one of the cutting edge topics calling for investigation. Several CODATA and National scientific communities involved in visualization would be eligible for this type of project.

An increasingly important area is bioinformatics as central in drug design, genome sequencing and the elucidation of protein shape and activity.

Academic/Industry Projects

French CODATA has decided to gradually expand its activities to include societal issues and has identified some fields of common interest to academics and to industry. To date, several areas have been identified:

- medical issues such as medical networks, patient/doctor relations, information transfer,
- certain types of industrial logistics,
- virtual university/E-learning (TIC problems),
- security in networking activities.

In all these prospective projects, the choice of exceptional scientists is essential, especially if one needs to use cutting edge methodologies in original and unusual applications.

In our exciting times, exploration, discovery and innovation lead faster than ever to rapid deployment of artifacts and commercial objects. The danger lies in emphasizing production in favor of prospective. This could be perilous when dealing with materials and instrumentation, for example.

Conclusion

In conclusion, may I recall that during the Renaissance, Leonardo da Vinci was both a scientist and an engineer. In previous centuries, most physicists and chemists (e.g. Priestly, Lavoisier, Haber, ...) were active as scientists and as entrepreneurs or industrialists.

This Knowledge Revolution is another Renaissance but more geared towards teams than towards individuals. The present exceptional combination of Emerging Knowledge and Efficient Assistance Tools provide today's scientists with the promise of vertiginous discoveries and innovation.

This article is dedicated to Professor Shizuo Fujiwara, a pioneer and promoter of the Knowledge Revolution.

抄録

データ、情報、知識 (DIK): 変化する社会における挑戦

ジャック・エミール デュボア (コデータ会長、パリ第7大学名誉教授)

貴情報知識学会誌への本稿で小生は、過去 50 年間に起こった変革を考える。そもそも科学技術 (ST) はこの期間にはじまったものではない。それ故、それ以前の、実証的で知的である人類の長い活動の成果である科学技術の特徴をまずあげてみよう。

19 世紀は情報革命の時代であり、それは知識革命でもあった。

ST で言えば、それはまず、ST 社会内で起こり、ついで世界全体のこととなった「情報流通」があげられる。これは「通信」の革命が原因である。

現在、ST 社会、一般社会がともに著しい変化状況にあるのがもう一つの特徴で、これはなお続くであろう。その状況のなかで、コデータがいかにあるか、その今後をナショナルレベル、国際レベルで考えて見たい。

科学と技術

まず過去 50 年間は物理社会 (physical world) の大きな変化が特徴である。良質でかつ大量のデータが得られるようになったが、問題も生まれ、その問題解決も錯綜にある。

18, 19 世紀は科学の専門分科が確立した時代であり、ガリレオ時代、決定論 (例えば数理解析の結論は確定可能とされる) 的時代であった。20 世紀の前半は、不確定性を基盤に置く量子力学の始まり、原子分子の微視世界測定、相対論による宇宙論がはじまった。これにつづく後半がまた重要な時期である。

コデータが創立されたのは 1960 年である。それはデータの質、データ集合の間の整合性などの問題を解決するためであり、また、情報交換のための標準化、数値データの κανονικός (正準的) セマンティクス (意味論) 的属性、コード化の文法などの課題が、個別性を持つデータを誰もが利用できるようにせねばならぬという要請にうながされての創立であった。

科学者、技術者から作る一次情報データは、科学の専門分野毎に作られたユニオンという効率のよいシステムで選別される。その際に要旨化作業までされるものも、それがない場合もある。その管理は“データ・情報・知識 (DIK)” システムの変革を呼ぶことになった。そして、この時期にコンピュータが登場した。

コデータはこの DIK の要請とコンピュータの導入の上に立ち、ソフトウェアの開発と、学際的な DIK の画像化、視覚化の活動を目指すこととなった。この間、新発明が進行した。レーザーのような光学の進歩、機械工学の進歩、装置的進歩などでデータはより精密になり、医学、物質科学、地理、地図学、天文学などにそれが及んだ。

この 50 年間の変化は、新しい情報処理の変化、ツールの進歩、学際性の増進、国際化に顕著である。

この変化を記述するには、これまでの十進法では不十分で、対数表示の SLS (Space Logarithmic Scale) がよいということをも 2000 年のバベノ (Baveno) のコデータ総会で提案した。(第 1 図)

どの分野でも、方法の精密化はデータの増大につながる。10 のマイナス 12 乗秒の物理化学測定は、1 秒あたり、 $\exp(\text{対数})$ の 10 の 12 乗の動的データを生むのである。このような世界のデータ処理は、これまでの我々の生活を規定してきた十進法の世界とはことなつた尺度、例えば、SLS 系が規定するような世界でおこなうべきことになるのである。ハッブル望遠鏡は 1 日に 1 ギガのデータを生み、それに対応するデータベースはテラバイト、人工衛星の Hypeccus データベースはペタバイトという具合である。

論理の表現問題も我々の現実社会がエコロジー、海洋関連の自然志向にあるように、同様に議論様式の変革が求められており、それはもとを探れば、精密科学ではじまり、単純系の体系をみることで始まった科学が、実は精密科学を逸脱した分野からの寄与で進歩しつつあるという現状によつていられる。以前この問題を古典的情報理論の枠内で考えることが提案されたことがあつた (北川敏男)。しかし、実際は、例えばファジー (fuzzy) 論理のような、多種の異なつた概念や手法のブレンドで進行するのであり、精密科学を基本とする現代 ST の内部構造の混交や、抽象化システムなどの、人間の異なつた活動や、文化が関わる問題として見るべきなのである。つまり、一つに偏した見方では対応出来ないのである。

この現状は、我々の生きる実社会が、任意性を増しつつあることに由来している。これは、自然が拡大化、任意性増大の方向にあることによるのであり、情報学会もこの大勢に対応をはかるべきである。例えば、不可視の“無限小”の議論は分子論的枠組みの問題でゲノム生化学で実際に扱われることになつた。“無限大”は「宇宙」の膨張にかかわつており、それは、始まり (origin) 人間の行動の拡大とも関わる。そこで“情報の概念”は、知識システム、認識論、DIK の相互関係に関わることになつるのであり、大、小は一体観に入るのである。これを第 2 図に示した。

かくて科学は単純系から多元世界に視線を向け変える。これは不可逆であり、自然の記述をその線にそつて進めねばならず、具体的には、それに適した数学的手段の整備が必要となる。

例えば、膨大な地球物理のデータは、近年おこりつつある、fractal structures, deterministic chaos, fuzzy logic, wavelet theory, strange attractor などを巡つて生まれた数学的モデル手法で処理されよう。(第 3 図)

1965 年に改善されたフーリエ変換でさえ、線形動力学に与えた恩恵は計り知れない。第 3 図には、近代 ST に及ぼした数学の寄与を示した。

ST の進歩が我々の概念や理論の限界のために制限を受けているのは否めない。情報革命は、この限界に由来する困難を取り除くところに特徴がある。

DIKW と通信

我々がもつ概念や理論の限界が複雑性の自然に対応出来ないことを以上述べた。それは自然、社会、科学の問題が複雑であるからであり、そこで、現在の所為にせよ、今後の企画にせよ、この対応をいかにするか意欲が起こるわけである。現代、新理論出現の間隔は短くなる一方である。そこで、分野間の理論の移転が重要になる。また、一束の新データにリンクして在来データ

を処理すると好結果が期待できる。それは、輸入種の培養の場合と同じである。現在、一分野で1日に得られるデータは、しばしば、その分野で過去数十年間に蓄積されたデータとより大きいことがある。以上の状況への対応は、まず、メタデータ群をつくりDIKをはかることである。それには、その作業に適した ontology を開発し、データを情報に変え得るようにしなければならない。それには、結果を知識化する能力(装置)を備えていることが必要で、移転、通信を可能にするものであってほしい。

現実問題として、DIK サイクル内相互間の連結では、D, I, K それぞれの内容の質とそれを処理する機能が問題である。例をあげよう。相関解析データは通常は特定の知識設計用に蓄積されたものが多い。それ故、そのようにして作られた古いデータ集を、仕立て直しをせずに新用途に利用するのは難しい。

同じことはDとIとの間でも起こる。とくにインターネット経由や迅速処理が必要な場合などではサーバーレベルのことで上記のような新旧混合の難が起こる。

「進歩」が並列、ないし、積み上げで連鎖すると、革命になり、真の発明につながることもある。よい例はレーザーであり、ニューラルネットワークである。これらは複雑な進化の結果生まれた。我々は、これは情報革命の一つと考える。

情報革命はいろいろな面の進歩にかかわっている。それ故、過去を承継する科学の基盤に対してもそうであって、例えば、現在、一次情報は、組織的に旧システムに入力される場合は、入力そのものが注射のように散発的であり、速度も遅いけれども、移転だと、インターネットのおかげで二次情報化の前の遙か早くに伝えられるのであり、科学基盤に対しても影響をするに至っているのである。

同様な趣旨で述べるのだが、新しいマルチメディアプログラムや、特定の研究やサービスを教える機能を持つ TIC(Technical Information and Communication) という新セクター(機関)が発足している。情報革命に教育が入ったわけである。

DIK サイクル内のデータ

データの定義は年月の間に变化した。その模様を第4図に示した。データの性質を分類すると、第一には、情報源の要素の内、数値や視覚情報のようなファクト(事項)的性質のものがあげられ、これは、計算機による意味解析で高次レベルに上げられる。第二は、科学データと普通言われるもので、処理ないし観測のレベルに応じてレベルが分類される。いわゆる、生ま、計算データの別である。第三は、観測データの類で、その数に対応して不完全性、不一致性、変動性などに関わっているとされる。これらは、第5図に示した知識管理システムのようにW(wisdom)レベルまで高める処理がされる。

ハード、ソフトデータ

技術的、概念的進歩によって、ソフトなデータ、ファジーなデータの世界に革命が起こりつつある。複雑系の処理システムは、人間の問題、社会の問題まで絡んできている。例えば、環境科学という持続性(sustainable)のある発展の問題。だから、このシステムの議論では、分子化学、薬学、さらには医学までが関わり、錯綜する処理システムの論理は、ファジー論理、ニューラル論理などまで必要とされるであろう。

そして、現実社会の問題を対象にするときは、より高い精密さをもとめるデータと知識のオペレーターとが複雑になり、答えは回帰的になり、また、単いつでなくなるのである。

異なる面を持つプロジェクトの解を得るにはチームワークが必要になることがある。あるいは、違った情報源、違った重みの源の交錯をはかる能力が必要となる。異分野の交流のチームには、異

種の言語による知識を活かす必要があり、また、最適化をはかる数学的ツールが使えるよう準備する必要がある。分類学、オントロジー、特異語、などの問題では、チームワークの上に、新基準と妥協の必要もあろう。

基本データやメタデータの構造化にあたってのそれらの標準化は、仮想現実実験室や多元ーエキスパートプロジェクトで仕事をするのに不可欠の条件である。いかなる DIK システムでも、通信ツールの平行的進歩や、遠隔ないし不均一なデータベース間の相互操作可能性がなかったら、働けない。データ転送やプロトコルが、サイト間の相互操作可能性で阻害されることは多い。

国際的、あるいは分野横断のデータは情報学の研究者に必要である。しかし、それは驚くべきことに、若い世代の教育にないのである。これは、我々の世代と仕事場が違っているのである。

我々の ST 世界が、現実社会に近づいていることは特記されることである。鳥瞰的視点がこれから必要である。

社会の複雑性の問題のなかで、知的所有権のことは重要である。

コデータの方向

コデータは専門分野の連携には努力してきた。そのために、データ収集と流通のタスクグループを作っている。

ナショナル活動

現下の動向

産学協同プロジェクト

適合度順検索システムの性能評価尺度

Performance measures for ranked output retrieval systems

相良 佳弘

大妻女子大学 (非常勤)

今日、WWW サーチエンジンのような適合度順検索システムが一般的になってきている。しかしこれらのシステムの性能評価を行う際に、従来の評価尺度では、このような検索システムの性能評価に必要な以下の四つの要件を満たしていないという問題点がある。

- 1) データベース中の全レコードの適合度判定を必要としない
- 2) ユーザ志向の適合度判定に基づいている
- 3) 適合度を多値的にとらえることができる
- 4) 順位付けの適切さを測ることができる

これまで広く用いられてきた精度、再現率は、これらの四つの要件をまったく満たしていない。また、順位付けを反映する評価尺度として考えられた正規化再現率やスライド比率も、四要件の全てを満たすには至っていない。本論文ではこの問題点を解決できる評価尺度として、順位相関分析、相関分析による評価尺度と、修正スライド比率を提案する。これらの尺度の妥当性を検討するため、実際に検索された23の検索事例を対象に実験を行った。その結果、これら三つの評価尺度は順位付けの適切さを十分に測ることができた。また、これら三つの尺度の特性も明らかになった。

- 1) 順位相関分析、相関分析に基づいた評価尺度は、網羅性が求められる事例に適している
- 2) 一方、修正スライド比率は、インターネット上での検索に見られるような、適合度が高い数件のレコードが得られればよい状況での評価に適している。

As ranked output retrieval systems such as WWW search engines become popular, development of reasonable measures to evaluate performance of such systems are increasingly needed. However, there are no satisfactory measures for the purpose since they lack the following four requirements.

- 1) Relevance Judgment of all the records in a database should be avoided.
- 2) User-oriented relevance judgment is needed.
- 3) Relevance measures should have multi-values.
- 4) Appropriateness of ranking outputs can be judged.

Neither traditional precision and recall nor normalized recall and sliding ratio that are used to evaluate ranked output retrieval system fulfill the above requirements. Therefore, measures based on a modified sliding ratio, rank correlation analysis and correlation analysis are proposed in this paper. To check the soundness of these measures, we experimented with 23-search examples and confirmed

that these three measures are adequate for evaluating such systems. In addition, two following characteristics are also found.

- 1) Measures based on rank correlation analysis and correlation analysis are suitable for searching when users need all relevant records.
- 2) A measure based on modified sliding ratio is suitable for searching when users need a few records with high degree of relevance.

1. はじめに

インターネットの普及に伴い、エンドユーザが大量の情報を直接検索する機会が増大している。こうした検索はサーチエンジンを用いて行われるが、そこでは一般的に適合度順検索システムが採用されている。このシステムは、従来のブール検索のように検索結果の集合を作るのではなく、蓄積レコードの順位付けを行っている。つまり適合性の概念を、ブール検索のように二値的にとらえるのではなく、多値的に判定し、その適合度順に検索結果が出力される。したがって、ブール検索で評価の対象となった検索集合における適合・非適合レコードの比率だけではなく、順位付けの妥当性がシステムの性能を測るうえで重要となる。そのため、新たに順位付けの妥当性を反映できる評価尺度を考える必要がある。

本研究では、インターネット上のサーチエンジンのような実際に運用されているシステムが、現実のユーザにとってどの程度有効であるかを評価する尺度が持つべき要件を考察する。そのうえで、順位付けの妥当性を考慮できるシステムの評価尺度について検討を行う。

2. 従来の評価尺度の問題点

2.1. 精度・再現率の問題点

クランフィールド実験^[1]以来、情報検索におけるシステムの評価尺度として、伝統的に精度や再現率が用いられてきた。しかし、精度や再現率には、いくつかの問題点が指摘されている。第一の問題点は、再現率を算出するためには、あらかじめ全蓄積レコードのそれぞれについて、適合か否かが判定されてい

なければならないことである。しかし、大規模なデータベースの場合、全レコードに対してこうした判定を行うことは労力の面から事実上不可能であるため、再現率を算出することはできない。TRECではこの問題を回避するために、大規模データベースを用いた実験を行う際、プーリング方式と呼ばれる手法を採用した。この手法は複数のシステム間で得られる適合レコードが、ある程度異なっていることを前提として、網羅的に適合レコードを決定しようとするものである。しかし、二つのシステムで全く異なる適合レコードが検索されるような状況下では、検索結果として出力されなかった適合レコードの存在が予想されるため、再現率を算出することは困難である。WWWのサーチエンジンでは、検索対象となるwebページを網羅的にデータベース化することは困難であり、このような問題が十分に起きると考えられる。

第二の問題点は、精度が検索集合における適合レコードの比率であるため、適合レコードが集合中の何番目に出力されたのかを表すことができない点である。仮に検索結果として20件出力した際に、5件の適合レコードが含まれていたとする。この時、この5件が1番目から5番目に出力されたとしても、16番目から20番目に出力されたとしても、精度は同じ値となる。従来のブール検索では検索結果が集合として得られるため、出力順は問題とならない。しかし適合度順検索システムでは、適合するレコードが何番目に出力されるかが非常に重要である。したがって、精度では順位付けの妥当性を測ることができず、適合度順検索システムの性能評価には不十分といわざるを得ない。

第三に、そうした検索システムでは、精度を算出する際に出力件数をどのように設定するかが問題となる。適合度順検索システムにおける出力件数は、ユーザが検索を終了するまでに参照した件数である。そのため、参照した件数によって精度の値は大きく影響を受けることになる。そこで、これまでに適合度順検索システムの評価においては、ある件数を出力した時点での精度が用いられてきた。また、再現率の算出が可能な実験用の小規模データベースを用いていた場合には、特定の再現率に対する精度や、精度と再現率との関係を示したグラフが多く用いられた。しかし、前述のようにサーチエンジンの場合には再現率の算出は困難であるため、精度しか用いることができない。この時、精度の値を算出するための適切な出力件数を決定することは容易ではなく、十分な評価が行えるのかが問題となる。

2.2. 主題志向の適合性概念

従来の情報検索システム研究では、主題志向の適合性概念に基づいて評価が行われてきた点も問題点として指摘できる。主題志向の適合性概念は、ある検索式が示す主題とあるレコードが示す主題とが一致するかを測るものである。従来の尺度は、あくまでも検索式に対して合致するレコードがどれだけ出力されたかを測るものである。この適合度判定では、ユーザと検索システムとの相互作用は一切考慮されていない。

しかし、実際に検索システムを利用するのは個々のユーザであり、ユーザを除外して性能評価を行ったとしてもそれは実態を反映したものとはいえない。ユーザは主題以外にも多様な要因から適合度判定を行っている。このようなユーザが行う適合度判定は、ユーザ志向の適合度判定と呼ばれている。現実の状況を反映させた評価尺度を考えるのであれば、実際にユーザにとってその検索システムが有用であるかを測ることが望ましい。そのた

めには、ユーザが情報要求を持つに至った背景などを含めた適合性概念に基づいて性能評価を行うことが必要である。

これまでに、実験用のデータベースを大規模化するなど、現実の検索が行われる環境に実験環境を近づける努力がなされてきている。しかし、実験環境からユーザに関わる要素が欠落しているため、現実とは異なる状況下での実験結果を得ているにすぎないといえよう。

3. 適合度順検索システムに適した評価尺度

3.1. 順位付けを考慮した評価尺度

前述のように、従来の尺度にはさまざまな問題があった。第一に、一般的に用いられている大規模なデータベースでは、再現率を算出するために必要な全記録レコードの適合度判定を行うことは不可能である。また第二に、個々のユーザが置かれている状況や情報要求の背景は、従来の評価尺度では反映されていなかった。実際に情報を利用するユーザを中心としたシステム評価を考えると、個々の状況における適合度判定を反映できることが望ましい。そのため、それぞれの検索事例において、ユーザが何件のレコードを参照し、それらのレコードの適合度をどのように判定したかを反映できることが求められる。第三に従来の精度・再現率では、二値的な適合度判定に基づいていた。しかし、適合度順検索では全てのレコードに順位付けがなされることから、適合度を多値的に捉えることが必要である。第四に、従来の評価尺度は検索結果における適合レコードと非適合レコードの比率に着目しており、どちらがより適合しているかといった順位付けは全く考慮されてこなかった。適合度順検索システムでは、検索結果の順位は性能評価に不可欠の要素である。そのため、このようなシステムの性能評価には、順位付けの妥当性を測ることのできる尺度が求められる。つまり適合度順検索システムの評

価尺度に必要な要件は、以下の四点である。

- ① データベース中の全レコードの適合度判定を必要としない
- ② ユーザ志向の適合度判定に基づいている
- ③ 適合度を多値的にとらえることができる
- ④ 順位付けの妥当性を測ることができる

検索されたレコードの出力順位に着目して評価を行う尺度として、これまでも正規化再現率、ESL(Expected Search Length)、スライド比率が考案されている。また、従来の精度に基づいて順位付けの妥当性を測ろうとした尺度として平均精度がある。以下では、これらの尺度が四要件をどの程度満たしているのかについて考察し、順位付けの妥当性を十分に測ることができるかを検討する。

3.1.1. 正規化再現率 [2]

正規化再現率は、適合レコードの実際の出力順と理想的な順位との差をとることにより、システムの性能を評価する試みであり、以下の式で算出される。

$$\text{正規化再現率} = 1 - \frac{\sum_{i=1}^p r_i - \sum_{i=1}^p i}{p(N-1)}$$

p 全適合レコード数

N 全レコード数

r_i i 番目の適合レコードの出力順位

i 理想的順位

たとえば表 1 に示すように、検索結果として得られたレコードのうち 2 番目と 5 番目に出力されたものが適合しており、その他は非適合であったとする。この時、2 番目と 5 番目のレコードが理想的な順位では 1 番目と 2 番目になると考えられる。正規化再現率は、実際の出力順位である 2、5 と理想的な順位である 1、2 の差から検索結果の質を評価する尺度である。仮にこの検索結果が 100 件のレコードを持つデータベースから得られた場合には 0.98 となる。また、この 2 件の適合レコードが 95 位と 98 位で得られたとすると、

値は 0.04 となる。つまり、適合レコードがより早い段階で出力されている場合、正規化再現率の値は高くなる。

表 1: 正規化再現率の算出例

出力順	判定	理想的な順位
1	非適合	
2	適合	1
3	非適合	
4	非適合	
5	適合	2

しかし正規化再現率は、データベース中の全てのレコードについて適合か否かが判定されている必要があり、従来の再現率と同様の問題点を持っている。つまり前述の①の要件を満たしていない。また、適合度の判定に、個々のユーザが置かれている状況は大きな影響を及ぼすと考えられる。しかし、正規化再現率は、現実のユーザが適合度の判定を行ったレコードだけではなく、その他のレコードの適合度判定も必要である。つまり、この尺度によるシステムの性能評価は、実際にユーザが検索を行う状況とは異なった環境下での適合度判定に基づいて行わざるを得ない。個々の検索事例におけるユーザの置かれた状況や、その時点におけるユーザの判定を反映することができず、要件②を満たすことができない。さらに、正規化再現率は適合度を多値的にとらえることなく、従来の再現率と同様に二値的な適合度判定に基づいているため、前述の例における 2 番目のレコードと 5 番目のレコードとではどちらがより適合しているかは、全く考慮されない。したがって、正規化再現率は出力順を用いて評価をおこなっているが、①から④の要件を満たしておらず、十分に順位付けを反映した評価尺度とはいえない。

なお、正規化再現率と同様に、実際の出力順と理想的な状態の順位との差から評価を行う尺度として正規化精度 [3] があげられる。正規化精度は、適合レコードが出力された順位の対数を総和した値と理想的な順位の対数を

総和した値との差を用いた尺度である。対数を用いているため、正規化再現率よりも高い順位で出力されたレコードに、より強い重みが付与される。しかし、正規化再現率と同様に四要件を満たすことはできず、順位付けの妥当性を測れる評価尺度とはいえない。

3.1.2. ESL^[4]

インターネット上の情報を検索する場合には、全適合レコードを出力するような高い網羅性を求める状況はほとんどないと考えられる。また、インターネット上に存在する膨大な情報の全てを確認することは、ユーザの負担の面から見ても困難である。したがって、上位で出力されるレコードが適合か否かがより重要である。この点から、ユーザが実際に参照したレコードを対象に、その部分の順位付けがどの程度妥当であるかを判定する尺度が求められている。

とくに上位の適合レコードのみに着目した尺度といえるのが、ESLである。ESLは、ユーザが検索結果に満足するまでに、どれだけの非適合レコードを参照したかで性能評価する尺度である。ESLでは、複数のレコードに同一の順位が付与されることを想定しているため、以下の式により計算される^[5]。

$$ESL = j + \frac{i \cdot s}{r + 1}$$

- j ユーザが満足する前の順位までに出現した非適合レコード数
- i ユーザが満足するレコードを得た順位に存在する非適合レコード数
- r ユーザが満足するレコードを得た順位に存在する適合レコード数
- s r 中の何件目でユーザが満足したか

この式では、複数のレコードが同順位と判定されることを想定している。たとえば、3件の適合レコードを得ることを条件とし、適合レコードが、2位、5位、7位で出力された事例を想定する。この時、6位までに判定されたレコードは全て1件ずつ出力されたが、7

位と判定されたレコードは3件で、適合、適合、非適合の順で出力されたとすると、ESLは約4.3となる。しかし、安形ら^[5]のように近年ESLを取り上げている研究では、全てのレコードに異なる順位が付与される状況を想定しているものが多い。この背景には、システムが複数のレコードを同順位と判定したとしても、実際には何らかの順番で出力が行われており、ユーザはこの出力順でレコードを判定していることがあげられる。つまり、同順位であることはユーザにとって意味がないが、一方で出力順はユーザの判定に影響を及ぼすことが考えられる。

そこで、出力順のように全てのレコードに異なる順位が付与されるとすると、2位、5位、7位で適合レコードが出力された場合、ESLの値は4.3ではなく4となる。この時、同じ条件で3位、4位、9位に適合レコードが出力された場合には、ESLの値は6となる。

ESLは、ユーザが実際に参照したレコードのみを用いて算出することが可能であり、データベース中の全適合レコード数が明らかでなくとも性能を評価することができ、四つの要件の①を満たしている。また、ユーザが適合か非適合かの判定をできるため、この点では、ユーザ志向の適合度判定に基づいているといえる。しかしESLでは、ユーザが満足する適合文献数をあらかじめ設定しておくことが必要となる。検索におけるユーザの満足については、様々な議論があるうえに^[6]、得られた情報に満足するために必要な件数は、情報要求ごとに変化するため、適切に設定することは困難である。検索を終える時点をユーザが任意に決定するのではなく、あらかじめ決めておくという点で、ユーザ志向とはいえない。つまり②の要件を完全に満たしているとはいえない。

また、ESLは適合度判定を適合と非適合というように二値的にとらえており、多値的にとらえるものではない。つまり、二つの適合文献間で、どちらの適合度がより高いかを評価することはできない。そのため、最も適合

しているとユーザに判定されたレコードが1番目に出力された事例と、3番目に出力された事例とは同評価となる。したがって、③および④の要件を満たすことはできない。近年、ESLを発展させて、ユーザが満足する件数を横軸としてグラフ化する試みが行われている^[5]。しかし、検索終了時点を適切に設定するのが困難であるという点を回避できるものの、四つの要件を満たすには至っておらず、順位付けの妥当性を十分に測れる評価尺度とはいえない。

3.1.3. スライド比率^[7]

適合度順検索では、適合度が高いレコードから順に出力されることが理想的である。そのため、適合度を多値的にとらえ、順位付けの妥当性を反映できる尺度が必要である。スライド比率は、個々のレコードに適合度の重みを付与することによって、こうしたことを試みた尺度であり、複数のレコードが同一の順位にならない場合には、以下の式によって算出される。

$$\text{スライド比率} = \frac{\sum_{i=1}^s w_i}{\sum_{i=1}^s W_i}$$

- s 評価の対象と指定された件数
- w 出力されたレコードに付与された適合度の重み
- W 理想的な順で出力されたレコードに付与された適合度の重み
- i 出力件数

スライド比率は、実際の出力順における重みの和と、理想的な出力順における重みの和の比である。たとえば表2のように、ユーザが、8件のレコードに対して0から4の五段階で適合度を判定した時に、検索例1に示すように適合度の重みが0、2、4、1、1、3、3、2であったとする。しかし、理想的な順位付けがなされている検索システムであれば、出力された8件のレコードの重みは4、3、3、2、

2、1、1、0の順になるはずである。検索例1の場合、3件出力時点での実際の重みの総和は6であるが、理想的な重みの総和は10である。したがって、この場合のスライド比率は0.6となる。

表 2: スライド比率の例

出力順	適合度の重み		理想的な重み
	検索例 1	検索例 2	
1	0	4	4
2	2	2	3
3	4	0	3
4	1	2	2
5	1	3	2
6	3	3	1
7	3	1	1
8	2	1	0

スライド比率では、データベース中の全レコードの適合文献数を知る必要がなく、任意の n 件の適合度が判定されていれば値を算出することができる。したがって、WWWサーチエンジン等のように大規模なデータベースにおいても、適用することが容易である。また、ユーザが判定した適合度の重みに基づいて性能を評価できるため、ユーザ志向の適合度判定に基づいているといえる。つまり、前述の①②③の要件を満たしている。

しかし、任意の n 件までに出力されたレコードの重みを合計し、その比を算出した値にすぎないため、対象となる検索結果の中での詳細な順位付けは反映できない。たとえば、表2における検索例1と検索例2では、明らかに検索例2の方が上位3件の順位付けは妥当であるといえる。しかし、双方とも得られた重みの和は等しいため、全く同じ値になってしまう。これは、順位付けの妥当性を十分に測ることができないことを示している。つまり、④の要件を満たしていないのである。

また、四つの要件の他に、ユーザが判定した n 件の一部しか評価の対象としない点も問題点として指摘できる。スライド比率を算出

する際には、ユーザが適合度を判定した n 件とは別に、評価の対象となる件数 s を設定しなければならない。この時、 s の値によってスライド比率は大きく影響を受け、 s の値が n に近づくとスライド比率の値も大きくなる。そして n 件全てを対象としてスライド比率を算出した場合、理想的な場合の重みの総和と実際の検索結果における重みの総和とが同じとなるため、スライド比率の値は必ず 1.00 となってしまう評価を行うことができない。このように、 s の値によってはその評価値が意味をなさなくなるため、スライド比率を用いて性能評価を行う際には、 s を適切に設定することが必要である。したがって、適合度を多值的にとらえることができるスライド比率にも問題点があり、適合度順検索システムの評価尺度としては適していない。

3.1.4. 平均精度

平均精度は、新たな適合レコードが出力されるたびに精度を算出し、その平均をとった値である。従来の精度に基づいた尺度であり、TREC などで幅広く用いられている。仮に、2位、5位、7位で適合文献が検索され、その他に適合文献がデータベース中に存在しないとすると、平均精度は 0.44 となる。また、同じデータベースを検索し 4位、7位、9位に適合レコードが得られた場合には、0.29 となる。つまり、この尺度ではより高い順位で適合レコードが得られる検索が高く評価される。平均精度は、従来の精度を用いつつ、順位付けの妥当性を測ることを試みた尺度である。

しかし平均精度を算出するためには、あらかじめすべての適合レコードの数が明らかになっている必要があり、四要件の①を満たしていない。また、正規化再現率と同様の理由で要件②も満たすことができない。平均精度では、適合度は従来の精度・再現率と同様に二値的な適合度判定に基づいており、前述の例における 2 番目のレコードと 5 番目のレコードとではどちらがより適合しているかは全く考慮されない。さらに、平均精度では適合度を二値的に捉えており、適合と判定されたレ

コードは全て同じ重みとして扱われ、多値的な適合度判定を反映していない。順位付けの妥当性を十分に測るためには、複数の適合レコードのうちどちらがより適合しているかを判定できることが必要であり、適合度を多值的に捉える必要がある。平均精度は、正規化再現率と同じく、出力順のより上位に適合レコードが出力されているか否かを判定できるが、適合度が高いレコードから出力されているかを判定できない。つまり、平均精度では、十分に順位付けの妥当性を測ることはできないといえる。

3.2. 順位付けの妥当性を測ることができる評価尺度

これまでに述べたように、さまざまな評価尺度が考案されているが、いずれの尺度も四つの要件を満たすものではない。そこで、これらの尺度が抱えていた問題点を解決し、実際の出力順序が理想的な状態にどれだけ近似しているかを数値化できる尺度が求められる。このような、比較を行う尺度として、順位相関分析・相関分析に基づく評価尺度と修正スライド比率の二つを提案する。

3.2.1. 順位相関分析・相関分析に基づく評価尺度

検索結果の評価を二つの変量間の関係ととらえるのであれば、順位相関分析や相関分析を用いることが考えられる。スピアマンの順位相関分析は、二つの変量の順位に何らかの関係が存在するかを測るものであり、以下の式で求められる。

$$\rho = 1 - \frac{6 \sum d^2}{n^3 - n}$$

d 理想的な順位と現実の順位の差
 n 対象となる件数

ここで、ユーザによって決定された理想的な出力順位と現実の出力順位との順位相関係数を評価尺度とすることを考える。もし、双方

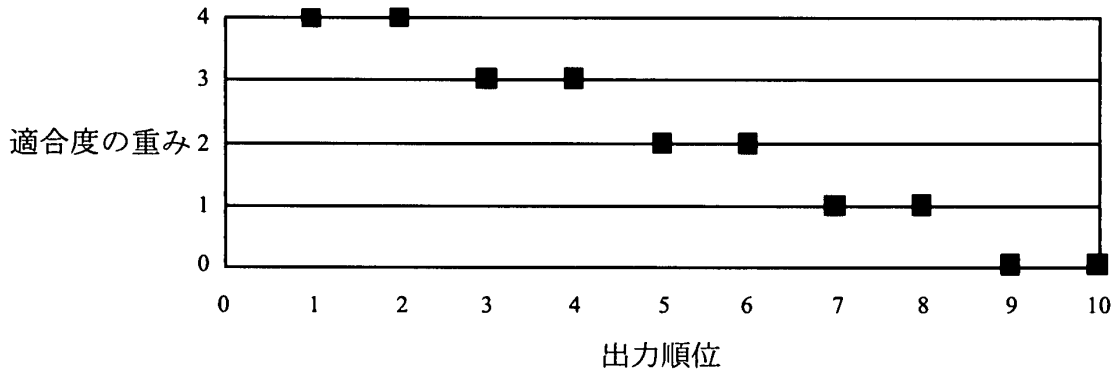


図 1: 適合度の重みと順位の散布度

の順位付けが完全に一致すれば値は 1.0 となり、理想的な順位付けが達成できたことになる。一方、値が小さくなる、あるいは負の値が得られたときは、実際の出力順が妥当ではなかったといえる。

順位相関分析を用いた場合、ユーザが参照しなかったレコードは評価の対象とはならないため、全適合レコードが明らかになっている必要はない。さらに、ユーザによって判定された順位を理想的な状態として比較しているため、ユーザ志向の適合度判定に基づいているといえる。また、個々のレコードの順位は、出力された全てのレコードの適合度を相対的に測ったものであるため、適合度を多値的にとらえていることになる。つまり前述の四つの要件を全て満たしており、適合度順検索システムの性能評価に適しているといえる。

しかし、順位相関分析を用いる場合には、出力された全てのレコードについて、理想的な順位が付与されていなければならない。出力レコード数が多くなった場合、その全ての順位を決定することは、被験者の負担を大きくする可能性がある。仮に 50 件の検索結果を出力した場合、被験者はこれらの全てに順位を付与しなければならない。このように、被験者の負担の面から、出力されたレコードに理想的な順位を付与することが困難な場合には、順位相関分析を用いることはあまり適切とはいえない。

これまでに適合度を多値的にとらえることを試みた研究では、被験者によって理想的な

順位を付与していくのではなく、適合度に何段階かの重みを付与していく手法が多くとられている。このように、適合度の重みが判定されている場合には、相関分析を用いて性能を評価することが考えられる。ここでの理想的な検索結果では、図 1 の散布図に示したように、適合度の重みと出力順位の間に関係が存在すると考えられる。この場合、出力順位と適合度の重みの間にマイナスの相関が得られ、分析の値が -1 に近いほど理想に近い順位付けであるといえる。相関分析を用いて順位付けの妥当性を測った場合も、順位相関分析と同様に、四つの要件の全てを満たしている。

3.2.2. 修正スライド比率

前述のスライド比率の問題点を解決し、より上位に適合レコードが出力されることを重視する新たな評価尺度を考えることができる。それをここでは、修正スライド比率と呼ぶ。この尺度は、単なる重みの総和の比ではなく、適合度の高いレコードがより上位に出力されていることも評価に反映させたものである。 n 件出力した場合の修正スライド比率は、以下の式で与えられる。

$$\text{修正スライド比率} = \frac{\sum_{i=1}^n \frac{w_i}{i}}{\sum_{i=1}^n \frac{W_i}{i}}$$

- n 出力件数
- w 出力されたレコードに付与された適合度の重み
- W 理想的な順で出力されたレコードに付与された適合度の重み
- i 出力件数

修正スライド比率では、スライド比率とは異なり完全に理想的な順位付けになった時のみ 1.00 の値をとり、より適合しているレコードが上位に出力された事例が高く評価される。また、スライド比率とは異なり、ユーザが適合度を判定したレコード全てを評価の対象とするので、対象件数 (s) を設定する必要はない。そのため、ユーザによって参照されたレコードすべてを対象として性能評価を行うことができる。たとえば、表 2 を例にとると、検索例 1 は 0.50、検索例 2 は 0.86 となり、検索例 2 の方が妥当な順位付けがなされていることが明らかになる。

相関分析や修正スライド比率を用いた評価では、 n 件までの出力に対して適合度の重みが明らかになっていれば評価の値を算出することができる。この場合評価は、被験者が理想的な検索結果を決定できるため、ユーザ志向の適合度判定に基づいているといえる。また、評価の対象となるのがユーザによって参照されたレコードであるため、全レコードの適合度判定を行う必要がない。このように、順位相関分析、相関分析、修正スライド比率による性能評価は、前述の適合度順検索システムの評価尺度に求められる四要件を全て満たしている。

4. 評価例

4.1. 評価例の概要

本論文では、これまでに述べたように順位付けの妥当性を評価できる尺度として、順位相関分析、相関分析、修正スライド比率をあげた。しかし、実際にユーザが WWW サーチャエンジンを検索するような状況下で、どの

評価尺度が適しているのか、また、それぞれの評価尺度がどのような特性を持つのかについては明らかにしていない。そこで、これらの尺度を用いて検索結果の評価を行った場合、それぞれの尺度間にどのような差異が現れるのかを、実際の 23 の検索事例をもとに検証する。

ここで用いたのは、相良 (1998) の実験で得られた 23 件の検索事例である。これらの事例は、書誌データベースを対象に適合度順検索を行ったものである。検索対象のデータベースの規模は WWW 環境におけるサーチエンジンと同様に大きく、また検索手法もサーチエンジンで一般的に用いられているキーワードの出現頻度情報に基づくものである。このことから、この 23 件の事例を用いることにより、四要件を満たす尺度が適合度順検索システムの性能評価を行えるかを十分に示すことができる。

また、この 23 件の事例では、検索システムが 30 位までに出力したレコードを対象に、それぞれに表 3 に示す基準に基づいた適合度の五段階評価および順位付けの評価が被験者自身によってなされている。そのため、順位付けの妥当性を評価尺度が反映できるかを確認するためには適したデータであると思われる。今回は、五段階のそれぞれに 0 から 4 の重みを与えた。

これら 23 件の検索事例の中で、修正スライド比率による評価が高い事例の評価値および 23 事例中の評価順位を示したものが表 4 である。一方、表 5 には、評価値が低い検索事例を示した。なお、相関分析の値は、順位相関分析との比較を行いやすくするため -1 を乗じている。

4.2. 高い評価が得られた事例

事例 22 は、修正スライド比率で最も高い評価値が得られた事例であるが、同時に順位相関分析、相関分析でも最も高い評価値が得られた。表 6 は、表 4 および表 5 でとりあげた

表 3: 五段階評価の基準

評価の基準	今回付与した重み
完全に情報要求に一致する	4
周辺のテーマも扱っているが、十分に情報要求を満たす	3
周辺のテーマが中心であるが、十分参考になる	2
周辺のテーマであり、あまり参考にならない	1
情報要求と全く一致しない	0

表 4: 修正スライド比率が高い事例

	事例 22		事例 23		事例 16	
	評価値	評価順位	評価値	評価順位	評価値	評価順位
修正スライド比率	0.966	1/23	0.900	2/23	0.892	3/23
順位相関	0.610	1/23	0.582	2/23	0.437	3/23
相関 (5 段階)	0.963	1/23	0.609	2/23	0.312	6/23

表 5: 修正スライド比率が低い事例

	事例 19		事例 11		事例 3	
	評価値	評価順位	評価値	評価順位	評価値	評価順位
修正スライド比率	0.323	21/23	0.290	22/23	0.262	23/23
順位相関	0.065	15/23	-0.086	14/23	0.058	16/23
相関 (5 段階)	0.104	15/23	0.206	12/23	-0.290	22/23

事例における理想的な順位と適合度の重みを、実際に出力された順に示している。事例 22 は 30 件の出力レコード中最初の 9 件が全て適合レコードであり、10 位以降の 21 レコードは全て適合度が低いと判定された事例であった。つまり、ほぼ理想的な検索結果が得られた場合であるといえる。また事例 23 も、適合度が高いと判定されたレコードが上位で出力されている。一方で、修正スライド比率の値が低くなった事例は、いずれも下位の出力でも適合度が高いレコードが出力されている。たとえば、事例 19 においては、適合度の重みが 4 と判定されたレコードが最初に出力されたのは 9 番目であり、順位付けが妥当に行えているとはいえない。事例 11 では、3 番目に適合度が高いと判定されたレコードが出力されているものの、下位の出力においても非適合レコードとともに適合レコードも出力されてい

る。これも順位付けが妥当とはいえない例である。

順位相関分析を用いた場合、修正スライド比率と同様に事例 22、事例 23、事例 16 の評価値が高くなった。相関分析でもほぼ同様の結果が得られており、事例 22、事例 23 に対して高い評価がなされた。表 4 に示すように評価値が高い場合には、修正スライド比率、順位相関分析、相関分析の三つの手法による評価は類似する傾向が見られた。このように、五段階評価に基づいた相関分析と個々のレコードの順位に基づいた順位相関分析の評価値とには、何らかの関係があるように思われる。もしそうならば、被験者の負担が重いと考えられる順位付けではなく、五段階評価に基づいた相関分析で十分に検索システムの性能評価が行えるといえる。

表 6: 評価例

出力順位	事例 22		事例 23		事例 16		事例 19		事例 11		事例 3	
	順位	重み	順位	重み	順位	重み	順位	重み	順位	重み	順位	重み
1	3	4	4	4	3	4	19	0	17	0	15	0
2	5	4	7	4	2	4	22	0	16	0	17	0
3	9	3	19	1	4	4	4	2	1	4	14	0
4	8	3	3	4	6	4	7	2	22	0	23	0
5	2	4	10	4	19	2	23	0	30	0	1	3
6	1	4	8	4	25	0	27	0	29	0	18	0
7	6	3	11	4	20	1	6	2	20	0	2	3
8	7	3	24	0	1	4	15	1	24	0	26	0
9	4	4	9	4	16	2	1	4	14	0	25	0
10	20	0	13	3	8	4	30	0	3	2	24	0
11	12	0	6	4	30	0	9	2	5	1	22	0
12	16	0	20	0	13	3	2	4	8	0	16	0
13	14	0	16	2	5	4	12	1	18	0	4	3
14	18	0	5	4	18	2	5	2	28	0	3	3
15	21	0	12	3	12	3	25	0	12	0	21	0
16	26	0	2	4	10	4	28	0	11	0	20	0
17	28	0	28	0	28	0	29	0	9	0	19	0
18	17	0	21	0	15	3	17	0	2	2	7	2
19	25	0	1	4	27	0	18	0	19	0	8	2
20	29	0	18	1	9	4	13	1	27	0	6	2
21	22	0	26	0	7	4	8	2	15	0	30	0
22	24	0	15	2	22	1	26	0	13	0	9	2
23	30	0	22	0	23	0	20	0	10	0	13	2
24	23	0	30	0	29	0	14	1	26	0	12	2
25	10	1	25	0	14	3	11	1	25	0	11	3
26	27	0	17	2	26	0	10	1	4	1	5	0
27	19	0	23	0	24	0	3	4	7	0	29	0
28	11	0	27	0	21	1	16	0	6	0	28	0
29	15	0	14	2	11	4	21	0	21	0	27	0
30	13	0	29	0	17	2	24	0	23	0	10	2

4.3. 低い評価が得られた事例

しかし評価が低くなった事例では、順位相関分析、相関分析と修正スライド比率とでは異なっていた。たとえば事例 12 は、修正スライド比率では 0.83 と高い評価が得られたが、順位相関分析では 0.02 と低い評価になった。

この事例では、上位 10 位までに適合度が 4 と評価されたレコードが 8 件中 4 件出力されたが、同時に適合度が 0 や 1 と判定されたものも 4 件出力されている。つまり適合度が高いレコードと低いレコードとが混在して出力された事例である。順位相関分析や相関分析で

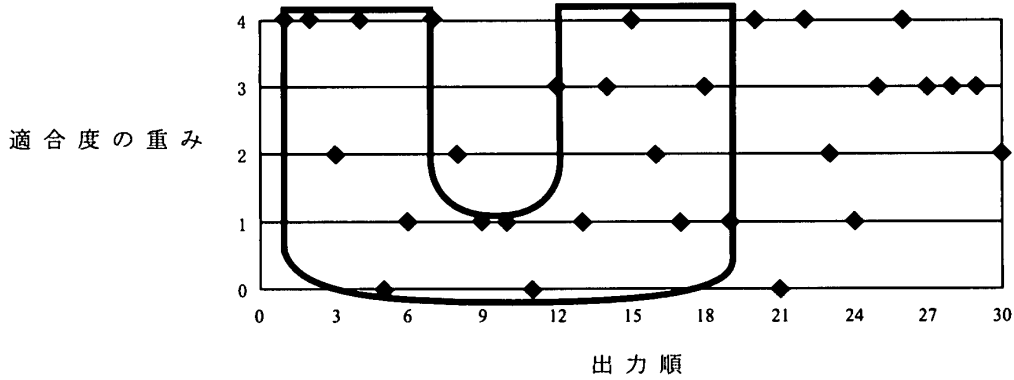


図 2: 事例 12 の散布図

は、上位に適合度が低いレコードも出力されているため、順位付けが妥当ではないとの評価がなされている。しかし、修正スライド比率では適合度が高いと判定されたレコードが上位に出力されているため、評価値が高くなっていると考えられる。

順位相関分析や相関分析と修正スライド比率の評価値が異なる場合、理想的な順位と実際の出力順との関係において、図 2 に示すような U 字状に似た関係が一部に見られた事例があった。事例 12 では、最初の数件の適合度が高く、その後 10 位前後で出力されたレコードの適合度は低いと判定されている。しかし、15 位に適合度が高いと判定されたレコードが再び出力されており、全体として図 1 の例で示したような関係にはなっていない。このように二変量間に直線的な関係が認められない場合には、相関分析では順位付けが一致していないとして評価の値は低くなる。一方、修正スライド比率は、全体の順位付けよりも、より早い段階に適合度が高いレコードが出力されることを重視している。そのため、事例 12 のような検索事例を比較的高く評価する。

理想的な順位と実際の出力順が直線的な関係でないことは、部分的には順位付けが妥当であったとしても、全体としては適切ではないことを示している。ユーザが参照した全てのレコードの順位付けの妥当性を厳密に測るためには、順位相関分析や相関分析の方が適していると思われる。適合度順検索では、ユーザが出力件数を任意に決定できる。したがっ

て、高い検索効率を得るためには、ユーザが適当な件数で検索を終了させることが必要になる。仮に、下位での出力されたものの中にも適合度が高いレコードが含まれていると考えられる場合、適合するレコードを全て出力したかをユーザが判断することは難しくなる。そのため、参照した全てのレコードに妥当な順位付けがなされていない場合、網羅性を求めるユーザにとっては、検索の終了を決定しにくいことになる。このように、網羅性を重視する検索においては、全体的な順位付けの妥当性を測ることのできる尺度の方が適している。

一方、適合度が高いレコードを数件求めるような検索では、上位に適合度が高いレコードが含まれているかが重要となる。このような状況では、修正スライド比率のように上位のレコードに重点をおいた評価が望ましい。今日広く行われている「インターネット上の情報を検索する」という状況では、ユーザは必ずしもデータベース中の全適合レコードが必要とは考えていない。そのため検索システムが上位で出力したレコードが重要であり、検索システムの性能評価には、順位付けの妥当性を反映することが求められる。修正スライド比率は、上位の出力に評価の比重をおいているため、インターネットを用いた情報利用の実情に即しているといえよう。

4.4. 従来の尺度による評価

また、順位相関分析、相関分析、修正スライド比率だけではなく、比較のために ESL、スライド比率、さらに伝統的な尺度である精度も算出した。なお、平均精度および正規化再現率はデータベース中の全適合レコード数が得られないため、ここでは取りあげない。これらの評価値を算出する際に、ESL では、ユーザが満足する適合文献数を 3 件とした。また、スライド比率は 10 件および 20 件を評価対象とし、精度は 15 位および 30 位まで出力した時点の値を示すこととした。ESL や精度のように適合度を二値的に判定する尺度では、五段階評価の重みが 2、3、4 とされたレコードを適合、0 または 1 と判定されたレコードを非適合とした。今回用いた検索事例のうち二つの事例では、適合文献が 3 件以上含まれていなかったため、ESL の値が算出できなかった。

この結果、ESL、スライド比率、従来の精度といった四要件を満たさない尺度の評価値は、修正スライド比率や相関分析とは異なった結果となることも明らかとなった。精度の場合、実際の検索結果から、あきらかに事例 22 と事例 19 では、事例 22 の方が妥当な順位付けであるにもかかわらず、30 件出力時点での精度は 0.3 と同じ値であった。これは、事例 19 では下位に適合レコードが出力されているため、全体としての出力レコード数に対する適合レコードの割合が事例 22 と同じになったからである。このように、ある件数での精度では順位付けの妥当性を測ることはできない。

ESL による評価は、精度と評価の傾向が類似していた。23 の検索事例を通じて、ESL の値は精度が高い事例で低く、精度が低下するにしたがって高くなる傾向が見られた。たとえば、15 件出力時点での精度による評価において、高い評価がなされた事例 6 や事例 16 には、ESL でも高い評価がなされた。同様に精度の値が低くなった事例 8、事例 11、事例 18 に対しても、ESL での評価が低くなっている。これは ESL が、ある一定の件数の適合文献を

得るまでに参照した非適合文献数であり、適合文献数を固定した精度といえるからである。このことから、ESL では精度と類似した評価となる。

スライド比率も、順位付けの妥当性が測れていない面が見られた。事例 8 では、上位 10 位までに、適合度が 3 と判定されたレコードが 2 件、2 と判定されたものが 8 件出力されていた。このうち適合度の重みが 3 のレコードは、6 位と 9 位であり順位付けがそれほど妥当とはいえない。そのため、修正スライド比率や相関分析による評価ではそれほど高い評価は与えられない。しかし、スライド比率ではある順位までの重みの総和であるため、このような詳細な順位付けの妥当性は反映されず、10 件時点でのスライド比率が 0.92 と高い評価がなされた。また、事例 16 と事例 11 とでは、明らかに事例 16 の方が優れているにもかかわらず、20 件時点でのスライド比率は表 7 に示すように事例 16 よりも事例 11 の方が高くなっている。この事例 11 や事例 19 では、20 位までの出力で算出したスライド比率の値が 10 位までのものと比較して大幅に高くなっている。前述のように、スライド比率は、評価対象件数を何件にするかによって、値が大きく変化する。これらの事例は、評価対象件数が多かったため、評価値が 1.00 に近づいた例と考えられる。このように、スライド比率では、評価対象件数を適切に設定しなければ順位付けの妥当性を測ることができない。しかし、何件を評価対象とすればよいかは一概に決定できないため、実際に応用することは困難だと思われる。

四つの要件を満たしていない尺度は、実際に検索された事例を用いて算出した事例においても、順位付けの妥当性を反映したとはいえない評価値が得られることが見られた。このことから、適合度順検索システムの評価を行う際には、四要件を満たした評価尺度を用いるべきである。

表 7: スライド比率

	事例 22	事例 23	事例 16	事例 19	事例 11	事例 3
スライド比率 10	0.970	0.800	0.725	0.440	0.600	0.240
スライド比率 20	0.970	0.900	0.800	0.700	0.900	0.621

5. 結論と今後の課題

5.1. 結論

検索システムとして適合度順検索システムが一般的になっているが、これまでに順位付けを反映できる評価尺度は十分に研究されてこなかった。順位付けの妥当性を評価に反映させる試みは、かなり以前から行われているが、一般的に用いられるには至っていない。この原因として、検索システム評価において順位付けの妥当性がこれまで重要視されてこなかったことが考えられる。しかし、適合度順検索システムにおいては、順位付けの優劣がシステムを評価するうえで重要であることはいうまでもない。そこで本研究では、大規模なデータベースを検索対象とする適合度順検索システムの評価尺度として、従来の評価尺度の問題点を踏まえ、順位相関分析、相関分析、修正スライド比率の三つの手法を示した。これらの手法は、今日の適合度順検索システムの評価尺度に求められる四つの要件を満たした尺度である。どのような評価尺度にも、適した状況や適さない状況は存在する。今回取りあげた中でも、順位相関分析、相関分析と修正スライド比率とでは、評価値が異なる事例も見られた。しかしこれらの尺度は、いずれも評価の値に順位付けの妥当性が反映されており、十分に適合度順検索システムの評価尺度として用いることができると考えられる。

また、五段階評価に基づいた相関分析と個々のレコードの順位に基づいた順位相関分析とでは、ほぼ同様な結果が得られる可能性が示された。検索実験を行う際に、出力された全てのレコードを適合度順に並べ替える作業は、被験者に大きな負荷を与える。一方、適合度

の重みを何段階かで判定することは、これまでの研究事例でも多く見られ、被験者への負荷はそれほどでもないと考えられる。適合度の重みを判定することによって、順位を全て判定した場合と同様の結果が得られるのであれば、五段階の重みに基づいた評価で十分であろう。

5.2. 今後の課題

本研究では、実際に評価例として取り上げた検索事例が 23 件と少なかったため、得られた結果が真に適切であるかは明らかになっていない。検索システムの評価は、ユーザの満足とも関連が深く多様な要素が影響を及ぼしている。しかし、どの要因がどのように影響を及ぼしているかは、十分に明らかにされているとはいえない。評価尺度の有効性を明らかにするためには、情報検索のあらゆる側面を把握し、それがどのように影響を及ぼすかを把握することが求められる。そこで、より多くの検索事例をもとに、その妥当性を検討することは意義があるといえる。その際、何件程度の事例を検証することによって評価尺度の妥当性を論じることができるといった部分の検討も重要である。

また、これらの三つの尺度を実際に用いるうえで、評価対象件数、つまり出力件数が適切に設定できるかどうかは注意する必要がある。これは、順位相関分析や相関分析は、出力件数が異なれば分析の範囲が変わるため、同じ検索要求であったとしても評価の値は異なるからである。同様に、修正スライド比率も、出力件数に応じて適合度の重みの総和が変化するため、評価値が異なる。そのため、出力件数を適切に設定することは重要である。

しかし、現実のユーザを想定すると、同じ情報要求で同じ検索システムを用いたとしても、常に一定の件数を出力するとは限らない。したがって、現実のユーザを想定してなんらかの件数を設定することは困難である。適合度順検索システムは、出力件数をユーザが任意に決定できる。そのため出力件数を前もって設定するよりも、ユーザが出力した件数を用いて評価を行った方が、現実在即しており妥当であろう。適合度順検索システムではユーザの探索行動が検索効率に大きな影響を及ぼすため、ユーザがどのようにして出力件数を決定しているかや、どのような要因から検索を終了するのかといった、適合度順検索システム特有の情報探索行動について検証していくことも今後の課題である。

6. 謝辞

この論文をまとめるにあたり、慶應義塾大学文学部の細野公男教授には多岐にわたりご助言をいただきました。ここに記し、感謝の意を表します。

(2001年10月10日受付)

(2002年5月27日採録)

参考文献

- [1] Cleverdon, C. W., Mills, J. "The testing of index language device". *Aslib proceedings*, Vol.15, No.4, p106-130 (1963)
- [2] Salton, G. "The evaluation of automatic retrieval procedures: Selected test results using the SMART system". *American Documentation*, vol.16, no.3, p209-222 (1965)
- [3] 正規化精度は以下の式で算出される。

$$\text{正規化精度} = 1 - \frac{\sum_{i=1}^p \log r_i - \sum_{i=1}^p \log i}{\log N!(N-p)!p!}$$

p 全適合レコード数

N 全レコード数

r_i i 番目の適合レコードの出力順位

i 理想的順位

- [4] Cooper, W. S. "Expected search length: A single measure of retrieval effectiveness based on the weak ordering action of retrieval system". *American Documentation*, Vol.19, No.1, p30-41 (1968)
- [5] 安形輝, 野末道子, 服部紀彦, 上田修一. "インターネットのサーチエンジンの評価尺度: ESL(Expected Search Length)を使った検索実験". *Library and Information Science*, No.37, p1-11 (1997)
- [6] 山本昭, 相良佳弘, 細野公男. "検索結果に対する利用者満足の概念". 三田図書館・情報学会研究大会発表論文集. p25-28 (1998)
- [7] Pollack, S.M. "Measures for the comparison of information retrieval systems". *American Documentation*, vol.19, no.4, p387-397 (1968)

著者紹介

相良 佳弘 (正会員)

慶應義塾大学大学院 文学研究科 後期博士課程 単位取得退学

修士 (図書館・情報学)

現在、大妻女子大学非常勤講師、産能短期大学兼任講師

情報検索における適合性概念や、適合度順検索システムにおける情報探索行動について研究している。

日本図書館情報学会、情報処理学会、三田図書館・情報学会、情報科学技術協会 会員

e-mail sagara@ceres.dti.ne.jp

図書館におけるレファレンス経験の知識化

A Knowledge Management System for the Questions & Answers at the Reference Counters of Libraries

岩澤 まり子 吉田 亜津美 大下 祥司 中西 陽子

図書館情報大学

利用者の情報要求をみたすことができるレファレンスサービスを図書館で提供するためには、図書館員のレファレンス能力の開発が必要である。本研究では、情報要求に適したレファレンスツールを選択できるようにするため、図書館におけるレファレンス経験を知識として蓄積する方法を提案し、レファレンスツール案内システムを試作した。

はじめにレファレンス質問と回答およびツールの概念化を行ない、上位概念を用いて主題内容を表現した。その結果、分野名、特徴名、情報項目名および索引名を使用して、回答としてのツールが選択できることがわかった。これらを組織化し、Web上で利用できるシステムを構築した。本システムにより、図書館においてレファレンス経験を知識として共有化することが可能となった。

A management system for the library service is proposed, where the reference experiences at the counters of libraries are accumulated in a knowledge base and the latter base works so as it guide users along the lines of their interests. The system is a guide system formed of reference tools of tree structure. In practice, the user selects any subject field according to his interest, i.e., that of the feature, the information item, the index of the reference under questions, or any tool the user wants to use. Hence, the proposed system will allow the users to share the reference experiences which have been accumulated in the libraries.

1. はじめに

図書館は、利用者が求めている情報を入手できるようにするために、情報支援および人的支援を行なっている。情報支援は提供可能な情報を充実をさせることによる支援であり、人的支援であるレファレンスサービスは、情報を求めている図書館利用者を、図書館員が情報を提供することにより援助することである。利用者の情報要求を満たすことができるサービスを提供するためには、図書館員にレファレンス能力が求められるが、レファレンス能力開発のためには、

- 二次資料についての知識を豊富にすること

- 組織としてノウハウを共有すること

が必要になる^[1]。図書館員が、二次資料についての知識を豊富にもつことができれば、二次資料を活用して、図書館内外の資料から情報要求に合った情報を探することができるようになる。すなわち、二次資料に関する知識獲得は、提供できる情報の範囲を拡げることに等しい。さらに、個人がもつノウハウを共有することができれば、経験の浅い図書館員でも情報要求を満たす情報を提供することが可能となる。したがって、レファレンス能力の開発は、図書館における情報支援および人的支援の両面から重要であると考えられる。

しかしレファレンス能力は、回答として提供

することができるレファレンスツールに関する知識と、同種の情報要求に過去どれだけ対応してきたかの経験とが重ね合わさって培われるため、個人にノウハウとして蓄積されていくことが多い。このノウハウを共有するために、技術を磨く勉強会が開催され^[2]、またレファレンス経験を記録する努力がなされている^[1]。さらに複数の機関の図書館員が協力して頻度の高いレファレンス経験であるFAQ(Frequently Asked/Answered Questions:よくある質問と回答)を選別し、レファレンス事例集等の冊子としてまとめる努力もなされている^{[3][4][5]}。

レファレンスの回答となるツールの探し方自体は、長澤雅男^[6]や大申夏身^{[7][8]}らによって紹介されているので、これらの書籍が参考になる。さらにFAQをまとめた事例集を参考にすれば、類似する質問に対しては比較的簡単に回答できると考えられる。しかし質問に組織名や地名等の固有名詞が使用されているため、どのFAQを参照すればよいか判断が難しいことが多い。また、冊子体として提供されることの多い事例集では、FAQの追加、修正、削除も容易ではない。

このため、情報要求の主題内容を選択するとその要求にあうツールを紹介するシステムが、Smith^[9]、阿部^[10]、土橋^[11]により提案されている。これらシステムではメニュー検索方式を採用しており、項目を選択していきただけで求めるツールにたどり着くことができる。簡単にツールを調べることができるが、過去のレファレンス経験を参照することはできない。

他にも仮想的なレファレンスルームをWeb上に設け、電子メールによる問い合わせや、簡単な調査(参考調査)を行なうことができるエレクトロニック・レファレンスという試みもなされている^[12]。具体例としては、1995年に提供が開始されたInternet Public LibraryのReference Centerがある^[13]。ここでは、FAQはデータベース化されているが、図書資料などの電子媒体以外のツールを調べることができない。

本研究では、図書館におけるレファレンスサービスを支援するために、レファレンス経験を概念化して蓄積し、知識として共有する方法を用いて、レファレンス経験の知識化を試みた。さらに電子化することにより可能となった、新たなレファレンス経験の収集、複数のレファレンス経験の積重ね、ネットワーク上での共有についても考察する。レファレンス経験の知識化には、事例集から収集したFAQおよび医薬分野の企業内図書館から提供されたQ&Aをモデルとして使用した。

2. レファレンス経験の知識化

レファレンス経験の知識化は、レファレンス質問と回答(Q&A)の概念化、組織化および共有化により行なった。Q&Aをそのまま蓄積すると、検索利用には固有名詞が必要になるとともに、類似のレファレンス質問にしか対応できない。一つのQ&Aに基づいて対応可能なレファレンス質問の範囲を拡げるために、Q&Aから質問とツールを抽出し、これらの概念化を行ない表現した。さらに概念化の結果を使用して組織化を行ない、知識ベースに蓄積し、Webを利用した共有化を試みた。

2.1. Q&Aの概念化

Q&Aの構成要素は、図1に示すとおり、質問と回答である。質問は、探している情報を表現する情報要求事項と、情報を探すために使用する既知事項から構成される。さらに情報要求事項は、探している情報の分野を表す分野名とその分野を特徴付ける特徴名、および探している情報そのものを表す情報項目名を用いて、その意味を表現することができる。既知事項は、探索の手がかりとして使用されるため、その意味は索引名として表現することができる。

同様に回答は、回答としてのツールと回答にともなう注記である補足事項から構成される。ツールには、ツールの内容を表す主題事

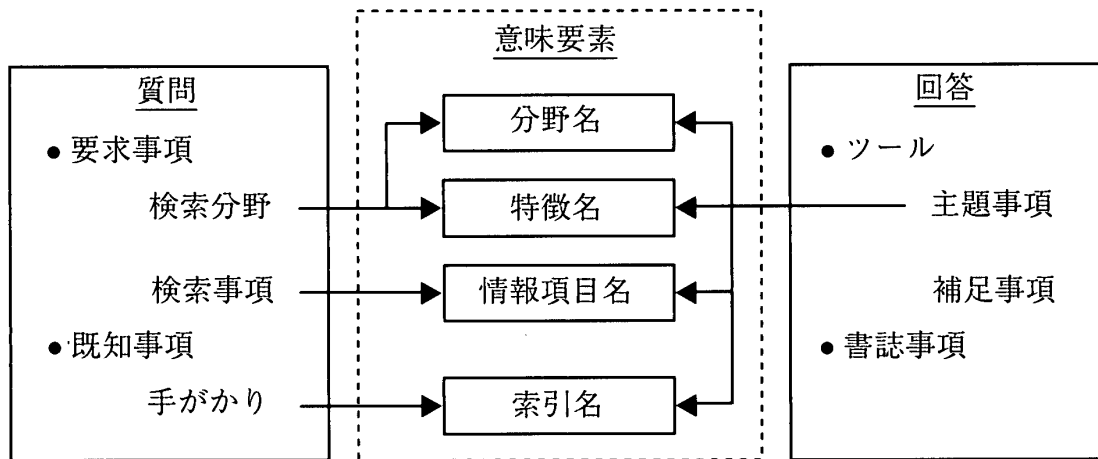


図 1: Q&A の構成

表 1: 主題分野の分類

分野名	内容
団体	企業や法人など団体に関するもの
人物	個人に関するもの
資料	本、雑誌、論文など資料に関するもの
ことば・文字	言葉の意味や読み方に関するもの
規格・法律	規格や法律に関するもの
統計	統計に関するもの
動物	動物に関するもの
化学物質	化学物質に関するもの
医薬品	医薬品に関するもの
事物・事象	「もの」、「こと」に関するもの（医薬品を除く）
地理	地理に関するもの
その他	上記のいずれにもあてはまらないもの

項とツールそのものを特定するための書誌事項がある。ツールの主題事項は、質問の意味要素を表すために使用した分野名、特徴名、情報項目名、索引名により表現することができる。したがって、Q&Aにおいて質問からツールが導き出される過程では、質問から情報要求事項と既知事項を表現する意味要素が抽出されて、ツールから抽出された主題事項を表現する意味要素との照合が行なわれていることになる。

意味要素の一つである分野名は、長澤による探索軸^[6]を基本として、医薬分野の図書館におけるQ&Aの特徴を考慮して、表1に示すとおり定めた。

特徴名は、日本標準産業分類^[14]で使用され

ているA~Nの分類記号に、「分類不明(Z)」を加えて使用した。日本標準産業分類では、企業の業種に加え公的機関の分類も可能であり、また新たな業種に使用できる分類記号が残っているため、特徴名の索引に利用した。

情報項目は、分野により大きく異なり、またツールによってその名称も異なっている。そのため、表1に示した分野に対応させて、情報項目を選定した。表2には、例として「団体」における情報項目名の概要を示す。

索引名は、ツールから情報を得るときにアクセスポイントとして使用される索引を示し、索引名もツールにより名称が異なっている。そのため、科学技術情報流通技術基準の索引作成^[15]を基にした統制語を索引名として使用した。

2.2. Q&A の組織化

Q&Aを概念化した結果、質問に対する回答としてのツールは、「探索事項」、「手がかり」、「探索分野」の組み合わせで決まることがわかった。ツール決定までの流れを、図2に示す。はじめに質問から、「探索事項」、「手がかり」、「探索分野」を決定する。探索事項を満たす情報項目をもつツールを探し、ツール候補ABCの集合を作る。この集合ABCから、既知事項を手がかりとして使用できるツールに絞り、ツール候補ABの集合を作る。さら

表 2: 「団体」分野における情報項目名の概要

ID	情報項目名	説明
1.00	概要	団体の概要についての情報
1.10	名称	団体の名称（正式名称、略称名）
1.11	連絡先	団体の連絡先（住所、電話番号、FAX 番号、e-mail）
1.12	ニュース	本、新聞、雑誌に掲載された団体に関するニュース
1.13	株主構成	団体の株主構成
1.14	会議	団体が行なった会議の情報（日時、場所、発行物、抄録）
1.15	一覧	団体に関するランキング情報（売上高、新卒採用、優良企業、自己資本率）、リスト
1.16	決算	団体の決算情報（発表日、決算情報、半期決算、年度決算、次期決算）
1.17	売上	団体の売上（半期売上、年間売上、次期売上）
1.18	資本金	団体の資本金
1.19	発行物	団体が発行する発行物（本、雑誌）
1.20	歴史	団体の歴史
1.21	名簿	団体の名簿
1.22	法律・条令	団体の法律や条令

に探索分野に基づいて主題分野をカバーするツールに絞り、最終的にツール候補 A の集合を得る。集合 A が、回答を導くツールとなる。そこで本研究では、図 2 に示したツール決定までの過程を、システム化することにした。探索分野は分野名と特徴名の意味要素により表現できるため、システム化においては、分野名、特徴名、情報項目名、索引名の四条件を使用することにした(図 3)。

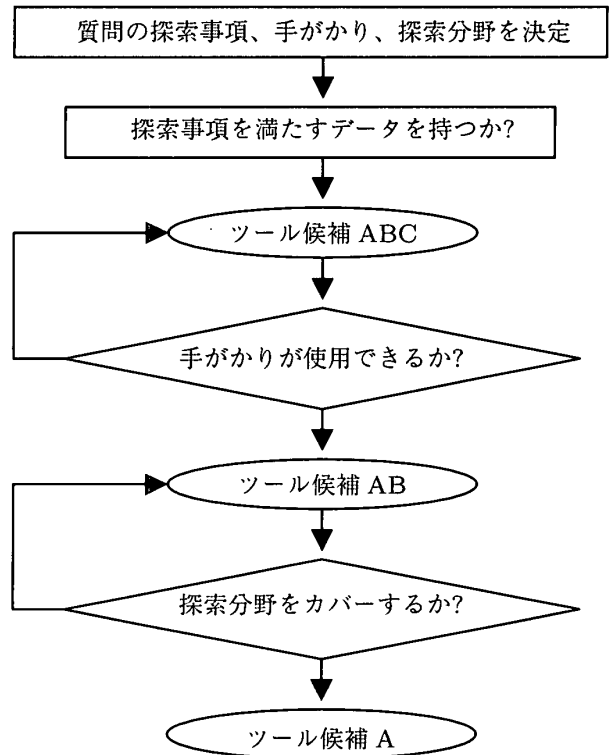


図 2: ツール決定までの流れ

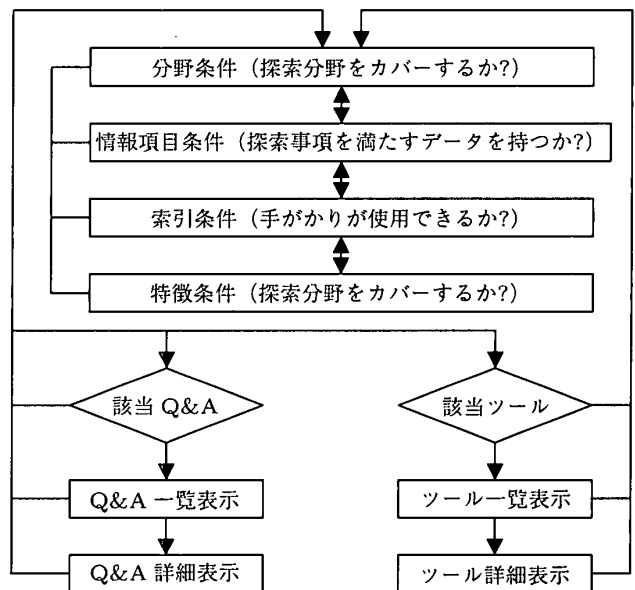


図 3: ツール決定モデル 1

はじめに、探索事項の主題分野である分類条件を、12 の分類項目の中から 1 つを選択することにより指定する。次に「探索事項」を情報項目条件として、「手がかり」を索引条件として、探索事項の主題分野の特徴を特徴条件として指定するが、いずれの条件も「指

定しない」を選択できるようにした。その結果、明確ではない条件が含まれるあいまい検索もできるようになった。検索結果としては、Q&A からツールを抽出して提示するが、四条件が一致した Q&A が蓄積されていた場合には、Q&A そのものも併せて提示することにした。

しかし、図3に示した組織化モデル1では、蓄積されている Q&A と同じ意味要素をもつ質問、および蓄積されているツールで答えられる範囲の質問にしか対応できない。つまり、一つの Q&A を利用して回答できる質問にしか対応できないことになる。

しかし情報項目条件が満たされないために、回答としてのツールが得られなかった場合でも、最初に得られたツールから新たな「手がかり」となる情報項目を見出すことができれば、次の四条件の組み合わせにより、最終回答としてのツールを見出すことが可能なこともある。このような複数の Q&A の積み重ねにより、最終回答としてのツールを得ることができるようにするため、図4に示す Q&A の積算を行なった。

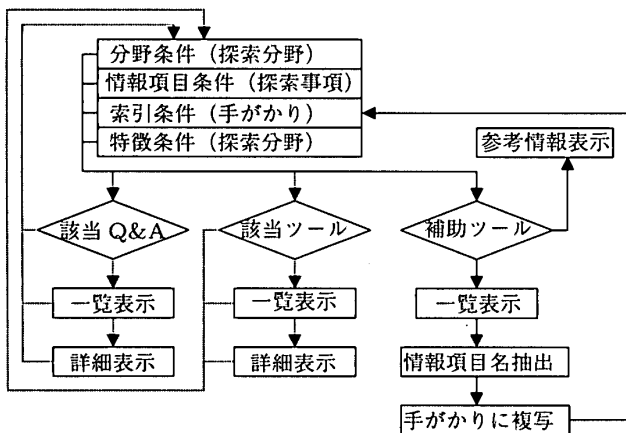


図4: Q&A の積算を含むツール決定モデル2

最初の分野名、特徴名、情報項目名、索引名の組み合わせにより補助ツールを得て、ここから情報項目名を抽出する。その情報項目名を手がかりである索引条件に複写し、二回目の分野条件、特徴条件、情報項目条件、新たな

索引条件の組み合わせを行なえば、最終回答としてのツールが得られることになる。例えば、「薬品の略称から開発企業を知りたい」という質問に対して、図3のモデル1では、同等の Q&A がなければ回答は得られない。しかし図4のモデル2では、システム内に「薬品の略称から正式名称を調べるツール」と「薬品の正式名称から開発企業を調べるツール」が Q&A として蓄積されていれば、これら二つを積み重ねることにより、ツールを求めることが可能である。したがって、本システムでは、積算が必要な二つのツールを回答として提示することができるモデル2を採用した。

Q&A を積算しても最終回答としてのツールが得られなかった場合は、手がかりを利用できるツールを参考情報として表示することにした。

2.3. Q&A の共有化

Q&A の共有化の概念を図5に示す。自ら解決できない情報要求が発生すると、図書館員への質問が行なわれ、その質問に対して情報要求を満たす情報が回答として図書館員から利用者へかえされる。このレファレンスサービスを繰り返すことにより、図書館員にはノウハウが蓄積されていく。この新たなノウハウを収集して対応可能なレファレンス質問の範囲を広げるために、Q&A 記録のためのフォーム作成し、電子メールを使用して新たな Q&A の収集を行なった。フォームには、Q&A を記

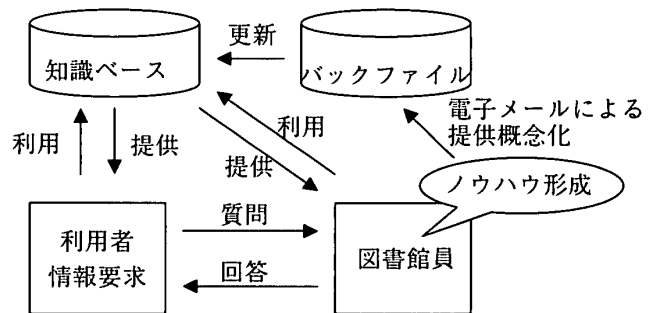


図5: Q&A の共有化

録するだけでなく、同時に質問とツールの概念化を行なうために、手順を記載した。概念化を複数の図書館員が行なうために生じるばらつきを少なくするために、一人の担当者による修正・確認を行なった。その結果を、知識として蓄積し、Web上で提供すれば、図書館員のレファレンスサービスを支援し、かつ利用者が自らを探すときの支援も可能となる。

3. レファレンスツール案内システム

3.1. システムの利用者

図書館員と一般利用者を、システムの利用者として想定した。このため、メニュー検索方式を採用するとともに、操作方法のヒントや注意メッセージを表示させることにより、使用経験の少ない利用者の支援を行なうようにした。

また図書館員は、本システムの利用者であるとともに、新たなQ&Aの提供者でもある。このため、図書館員がQ&Aの概念化を行なうことができるようにするため、Q&A収集のためのフォームを作成し、提供した。フォームには、概念化のばらつきを少なくするために、Q&Aを構成する意味要素ごとに説明をつけ、手順を示した。

3.2. Q&Aの情報

日本薬学図書館協議会発行のレファレンス事例集 — 薬学図書館編 — を、初期の情報源として使用した^[4]。この事例集には、13種類に分類(医薬、副作用、毒性など)された150件のFAQが収録されている。また専門図書館がまとめた事例集^[3]および公共図書館がまとめた事例集^[5]も、適宜、情報源として使用した。これらの事例集ではFAQとして、事例番号、質問文および回答文、使用されたツールが記載されている。また、一部のツールには編者のコメントが掲載されている。しかしこれらの事例集には、過去のFAQが蓄積されて

いるため、事例集の発行から時間が経過するにしたがいツールも古くなり、新しい内容のレファレンス質問には対応できなくなる。新しい質問に対応できるようにするため、医薬関連企業の情報担当者から、質問がくり返されられると思われるQ&Aを選択し、概念化を行なったのち、提供を受けた。

このようにして収集したQ&Aに事例番号を付与して、質問文と回答文、使用されたツール名、コメント、および事例集を使用した場合にはその書誌事項を情報として蓄積した。

3.3. ツールに対する情報

回答として使用されたツールに関する情報は、実際にツールを確認し、図書や雑誌の場合は、表紙、奥付、目次などから収集した。現物を確認できないものや、利用制限のあるツールについては、他の資料を参考にして確認を行なった。またツールとして示されているWebページはすべて確認し、利用できるものについては確認した日時を収集し、確認できなかったものには「現在は使用できない」という情報を記録し、過去の探索履歴は保存した。

ツールの情報としては、書誌事項の他に、情報項目名、索引名、ツールの提供形態、対象地域、利用条件等を収集し蓄積した。ツールの提供形態は、『日本目録規則1987年改訂版』^[16]を基にし、「第I部記述」の「第1章記述総則」を除く12章の項目名に、インターネットやオンラインデータベースなどを分類するための“14 ネットワーク情報”，他機関に問い合わせる場合なども考慮した“15 その他”の項目を追加して、使用した。

3.4. アプリケーション

システムは、webサーバ、サーブレットエンジン、リレーショナルデータベースを連携させるwebアプリケーションを用いて構築した。本研究ではMicrosoft Windows 2000 Server

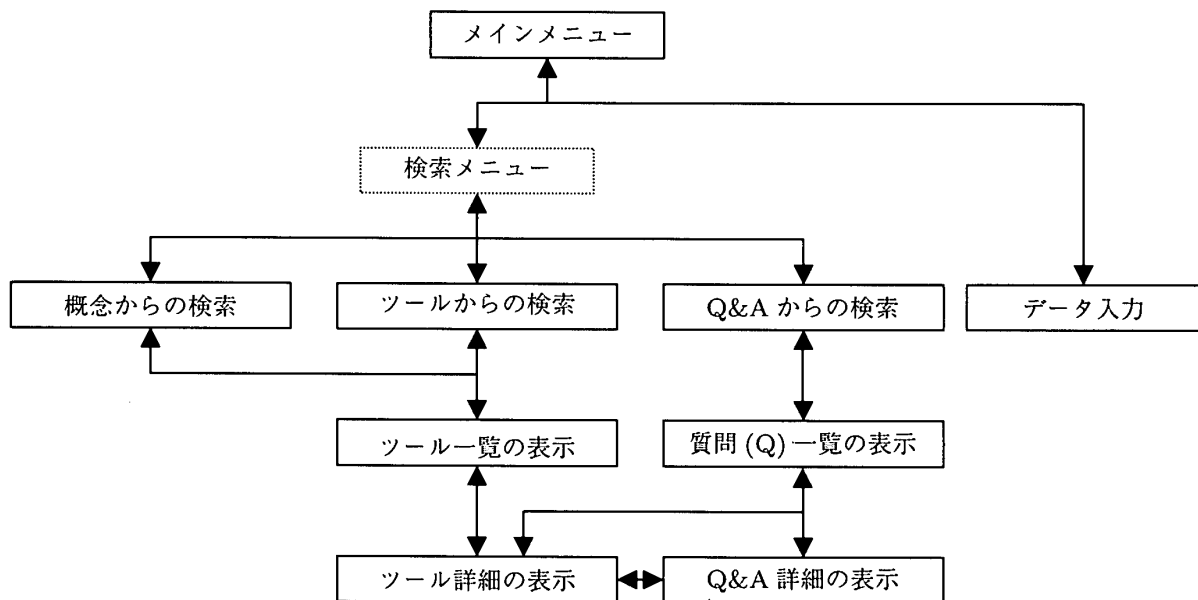


図 6: システムの画面構成

上で IBM WebSphere Application Server3.5 を稼動させた。データベースは Microsoft Access 2002 を使用した。

クライアント側は、サーバと通信できる環境で web ブラウザが使用できることを条件とする。本システムでは、画面遷移のコントロールに JavaScript を、画面レイアウトに Cascading Style Sheets(CSS) を使用した。

3.5. システム

本システムの画面構成を図 6 に示す。検索は、「概念からの検索」、「ツールからの検索」、「Q&A からの検索」を提供し、さらに新たな Q&A を入力するメニューを作成した。データ入力画面以外の各画面は自由に移動することができる。

「概念からの検索」では、概念化によりツールを分類名、特徴名、情報項目名、索引名の意味要素を使用して表現・蓄積することにより、質問からツールを検索できるようになった。しかしツールから、現在の調査と似た探索過程を持つ質問を特定することができない。そこでレファレンス質問を、ツールと同じ意味要素を使用して表現した。その結果、現在の調査と似た探索過程を持つ質問を特定する

ことができるようになり、これを「ツールからの検索」とした。さらに、過去の Q&A で使用されていないツールでも候補ツールとして、意味要素を使用することにより検索できるようになった。

しかし、過去の質問で使用されていないツールも検索されるため、過去の Q&A をシステム上に再現できなくなる。そのため、意味要素を介さないデータ構造を含めることとした。すなわち、検索の際には意味要素を介して質問からツール候補を提示し、その探索過程と同じ意味要素を持つ Q&A を提示する。提示した Q&A の詳細を参照するときには、意味要素を介さずに過去の質問に使用されたツールのみを表示する。

「Q&A からの検索」では、蓄積されている Q&A を一覧表示するブラウジング機能を提供した。

ブラウザを起動し、所定の URL を入力すると、メインメニュー画面 (index.html) が表示される (図 7)。

「概念からの検索」をクリックすると、図 4 に示した処理にしたがい検索を行なう。はじめに分類条件を入力する。ここでは、探索事項の主題分野を分析して作成した 12 の分類項目の中から 1 つを選択する (図 8)。さらに

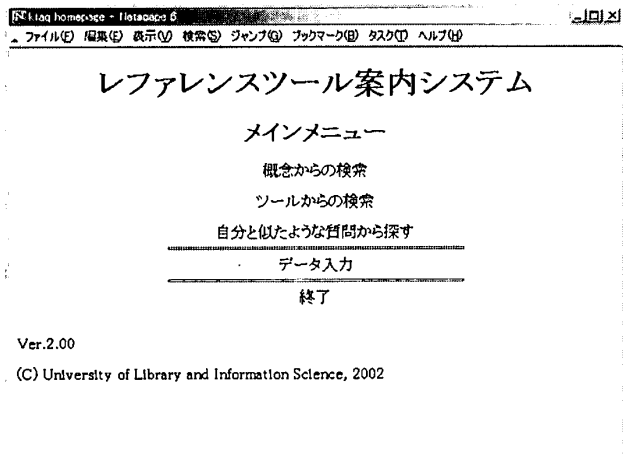


図 7: メインメニュー

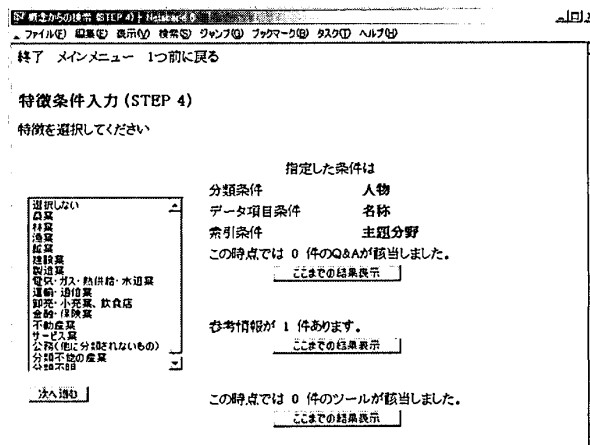


図 9: 概念からの検索画面-2

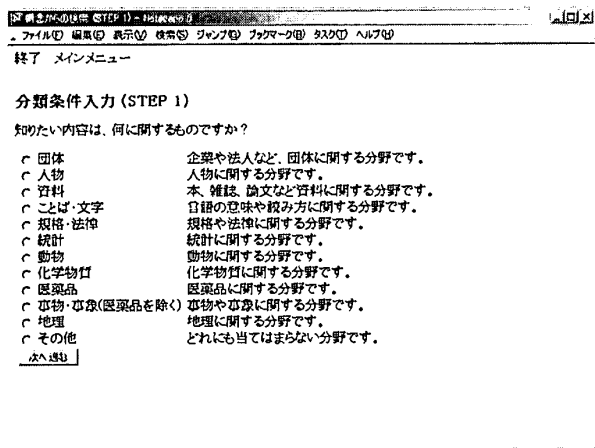


図 8: 概念からの検索画面-1

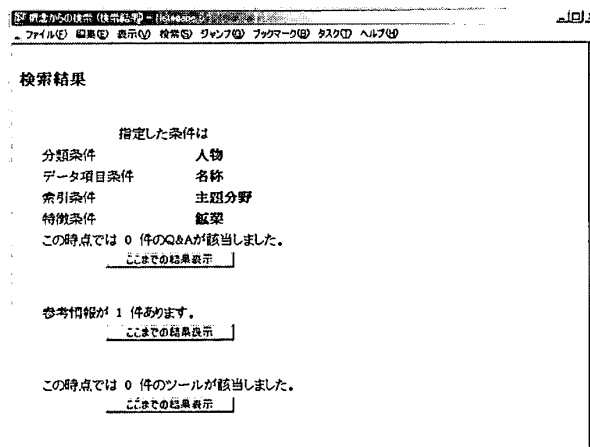


図 10: 概念からの検索結果画面

情報項目、索引条件、特徴条件を順に指定するが、各条件は「選択しない」を選び、指定しないことを選択できる (図 9)。結果としては、該当する Q&A およびツールの件数、参考情報の件数が表示される (図 10)。

「概念からの検索」から、該当するツール一覧を表示させた場合や「ツールからの検索」を行なった場合には、該当するツール一覧が表示され、さらにツールの詳細を表示させることができる。「概念からの検索」を行ないツール一覧の表示を介して Q&A 一覧を表示させた場合や、メインメニューから「Q&A からの検索」を選択した場合、条件に適合する Q&A 一覧が表示される (図 11)。ここから質問文のリンクをクリックすると、選択した Q&A の詳細を参照できる。

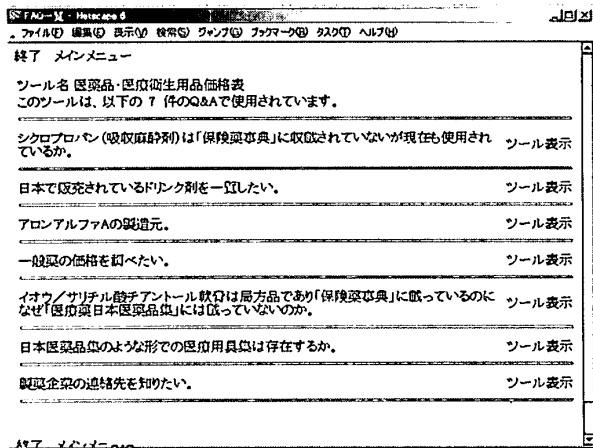


図 11: ツール一覧の表示画面

4. 考察

本研究では、情報要求者の支援を行なうことを目的に、図書館におけるレファレンス経験を活用したレファレンスツール案内システムを作成した。作成するシステムには、利用目的を考慮して次の機能を備えた。

- 1) ツールを検索する機能
- 2) 利用者と同じ探索過程をもつ Q&A を提示する機能
- 3) 関連のあるツールと Q&A を検索する機能
- 4) マウス操作により検索ができる機能
- 5) ツール名からそのツールについての情報を得られる機能
- 6) Q&A の新規登録、ツール情報の追加、更新をする機能

4.1. 汎用性

回答は「ツール名」という形で一般化されているが、質問は固有名詞をそのまま用いて蓄積し、検索に用いているシステムが多い。そのため、固有名詞を知らなければツールを探すことはできず、また一つの Q&A の経験から対応できる質問は限られていた。しかし本研究では、Q&A の質問内容およびツールを、分類名、情報項目名、索引名、特徴名の同じ意味要素を用いて表現したシステムを作成した。その結果、ツールが検索できるだけでは

なく、過去の類似した Q&A もあわせて検索・参照できるようになった。

レファレンス質問が情報戦略にかかわる内容である場合には、情報要求が明確にされないため、すべての意味要素をはっきりさせることができないこともある。このようなあいまいな質問でも、各意味要素の選択肢の一つである「選択しない」を選ぶことにより、ツールおよび Q&A を検索することができるようになった。

さらに二つの Q&A を積算が可能となったため、蓄積されているレファレンス経験が限られていても、対応可能な質問の範囲を拡大することができた。

4.2. 操作性

検索に関わる操作性を高めて容易に検索できるようにするため、システムの起動から終了までをマウスの操作だけで行なうことができるシステムを作成した。検索語の入力も、メニュー検索方式を採用することにより、画面に表示された索引語をマウスで選択するだけで、情報要求に合うツール候補を検索できるようになった。その結果、ツールについて詳しい図書館員と共に、ツールについてあまり知識のない一般利用者や本システムを使い慣れない利用者も、ツールの選択に本システムを利用できるようになった。

このように本システムは、検索の知識があまりなくても検索ができる画面構成を持ち、指定した条件と類似する過去のレファレンス経験を参照できる機能を備えているため、誰でもがツールを探ることができるようになった。しかし検索における操作性は優れている反面、質問とツールを細かく分析する必要があるため、過去に提案されたシステムに比べてデータの修正・追加における負担が大きい、という欠点がある。

4.3. 共有化

本システムは、Web アプリケーションを用いて作成し、ネットワーク利用を可能とした。また、日本語をサポートしない Web ブラウザを用いた利用も可能であるため、医薬分野の企業内図書館のように世界に図書館員と利用者が分布している図書館においても、レファレンス経験の共有が可能となり、利便性が高まった。

また図書館員と利用者との間のレファレンス経験の共有化においても、本システムは有効であると考えられる。すなわち、一般利用者が本システムを利用すると、次の効果が期待できる。

- 図書館員のノウハウであったレファレンス技術を共有することができる
- 図書館員に相談しなくても、ツールを探ることができる
- 図書以外のツールも含めて、適するツールを選ぶことができる
- 利用館以外の図書館にあるツールも、検索の対象とすることができる
- Q&A の履歴から、新たなツールの活用法を見出すことができる

また、図書館および図書館員には、次の効果が期待できる。

- 本システムを用いて一般利用者が自由にツールを探ることができるようになるため、レファレンス業務の軽減をはかることができる
- 利用者が解決できない高度な質問について、図書館員が十分な時間を割くことが可能となる
- 保有しないツールに対する問い合わせの頻度をもとにして、新たに備えるツールの選定ができる

4.4. 情報収集

システムに取り込むツールは、図書やオンラインデータベースなど、その形態を問わず

対象として、適するツールを選択できるようにした。情報源としては事例集を使用した³が、事例集のみでは過去の FAQ の収集に限られてしまう問題がある。発行から時間が経過するとともに、FAQ ではなくなることも多い。システムの有用性を保持するためには、蓄積データの確認・更新が必要となるが、同時に新たなレファレンス経験を継続して収集することも必要である。そのため Q&A 入力フォームを作成し、将来 FAQ になる可能性がある⁴と図書館員が判断した新たな Q&A を収集した。Q&A は、医薬分野の企業内図書館に勤務する図書館員が選択し、概念化を行なったものを使用した。

医薬分野の企業内図書館に協力を求めたため、企業内図書館という利害関係も起こりうる組織間の知識共有もテーマとなった。Q&A を概念化して表現・蓄積することにより、質問内容が一般化されて表現されるため、各社に不利益が生じない形での共有・協力体制のありかたが提案できたと考える。

5. おわりに

本研究では、レファレンスツール案内システムモデルを提案したが、本システムを実際に運用するためには、解決すべき問題が多数残されている。本システムでは汎用性を高めるために Q&A とツールの概念化を行なったが、医薬分野に特化したため、各属性の属性値の深さにばらつきが生じた。広い分野を大使用とするためには、より一般的な索引体系を考えなければならない。

また検索条件として四意味要素のみを使用した⁵が、蓄積されたデータが増えたときには、ツールの“利用の可否”や“提供形態”などの補足事項として蓄積されている項目を用いて、使用できるツールだけを検索できる機能も必要となる。またシステム内に、“利用の可否”などのローカルな情報を蓄積する方法の提供も必要となる。

インターネットが発達した現在、自ら Web

を検索して情報要求を満たすようになり、図書館に寄せられるレファレンス質問は、Webの発達と共に変化していると考えられる。今後、Q&Aの提供者として図書館員に一般利用者を加え、またツールとしてWebページを加えることにより、図書館員と利用者との連携を深め、多くの有用なレファレンス経験を知識として活用する方法を検討していきたい。

謝辞

土谷久氏をはじめとする医薬分野の企業内図書館の情報担当者の皆様に、企業内図書館におけるQ&Aを提供して頂きました。ご多用にもかかわらず協力頂きました皆様に、深く感謝致します。

なお、本研究には、図書館情報大学特別研究の補助を頂きました。

参考文献

- [1] 山崎久道. 専門図書館経営論 — 情報と企業の視点から. 東京, 日外アソシエーツ, 1999, p.169.
- [2] 真銅解子. 同報メールで作った『レファレンス事例100』. 図書館雑誌. Vol.93, no.8, 1999.8, p.626-627.
- [3] 三井業際研究所十日会ライブラリアン連絡会編. レファレンス事例100 — 企業ライブラリーに寄せられるこんな質問あんな質問 —. 1998, 121p.
- [4] 日本薬学図書館協議会関東地区研究会編. レファレンス事例集 — 薬学図書館編 —. 相模湖町, 日本薬学協議会, 1998, 80p.
- [5] 東京都立中央図書館資料部参考課事例集担当者会編. 東京都立中央図書館レファレンス事例集. 東京, 東京都立中央図書館, 1994, 152p.
- [6] 長澤雅男. 情報と文献の探索. 第3版, 東京, 丸善, 1994, 337p.
- [7] 大串夏身. チャート式文献・情報アクセスガイド. 東京, 青弓社, 1992, 228p.

- [8] 大串夏身. インターネット時代の情報検索術. 東京, 青弓社, 1997, 175p.
- [9] Karen F. Smith, Robot at the Reference Desk?, College & Research Libraries, Vol.47, no.5, 1986, p.486-490.
- [10] 阿部悦子. 参考業務システムSERENDIPの開発. 情報の科学と技術. Vol. 47, no.1, 1997, p.35-42.
- [11] 土橋喜. ハイパーメディアによるデータベースディレクトリと分類体系のシステム化レファレンス支援システムのための構想と試作. 大学図書館研究, no.45, 1994, p.8-15.
- [12] 大塚敏夫. エレクトロニック・レファレンス. 現代の図書館. Vol.38, no.1, 2000, p.44-53.
- [13] Internet Public Library. Reference Center. (参照 2002/5/24)
〈URL: <http://www.upl.org/ref/>〉
- [14] 総務省統計局統計センター. 日本標準産業分類分類項目表. 平成5年10月改訂. (参照 2002/5/24)
〈URL: <http://www.stat.go.jp/info/seido/11.htm>〉
- [15] 科学技術庁科学技術振興局. SIST13 索引作成. 1998年版. 東京, 科学技術振興事業団, (SISTハンドブック科学技術情報流通技術基準) (参照 2002/5/24)
〈URL: <http://www.jst.go.jp/SIST/sist13/sist13-m.htm>〉
- [16] 日本図書館協会目録委員会. 日本目録規則1987年版. 改訂版, 東京, 日本図書館協会, 1994, p.19-235.

(2002年5月28日受付)

(2002年6月21日採録)

著者紹介

岩澤 まり子 (正会員)

理学博士

図書館情報大学 図書館情報学部 助教授

Email: miwasawa@ulis.ac.jp

吉田 亜津美 (非会員)

図書館情報大学 図書館情報学部

(現在 株式会社ジー・サーチ)

大下 祥司 (非会員)

図書館情報大学 図書館情報学部

(現在 有限会社志群)

中西 陽子 (学生会員)

図書館情報大学大学院 情報メディア研究科 博士課程在学中

Bibliometric Analysis on Japanese Advertising Activities after the End of the World War II

Part I: Phenomenological features extracted from one specialized journal on advertisement

Hideya Azuma* Junichi Sato[†] Shuichi Iwata[‡]

* SENDENKAIGI Co., Ltd., President

† International Center for Research and Development of Metatechnica Technology, President

‡ The University of Tokyo, RACE, Professor

Advertisement could be regarded as a technology of word and language in social and economic domain, which makes economic activities closely related with life and social basic system in the modern world. This report gives the result of the bibliometric analysis on the words collected from all the titles of monthly SENDENKAIGI, the only one Japanese journal for specialists in advertisement, examining its 613 issues from 1954 to 2000. It pursued phenomenologically the change of Japanese economy, industry and society, and their relations have been investigated through the usage frequency of words for advertisement by decade from 1950's to 1990's, that is, from the ruins of the Second World War to the present position of one of the leading countries in the world.

By studying the meaning and role of advertisement as wide as possible, even on such situations as human, society and environment besides its original role for market communication, it was phenomenologically made clear that the advertisement not only plays so important roles for suppliers to introduce their products to consumers, but also such extending roles as methods of communication to disseminate the situations of economic development and the social state to give to people the perspectives for the following decade hereafter on the economic and social direction, business concepts, and further environmental problems and international situation.

現代社会で生活の基本システムである経済活動の社会連関において、言語を道具とした社会的、経済的活動である広告に着目し、日本唯一の広告専門誌「宣伝会議」の1954年から2000年の47年間613冊の全誌における掲載記事の全タイトルに用いられた言葉の出現頻度について、ビブリオメトリー分析を実施した。50年代から90年代にわたる10年期毎に、第2次大戦後の日本の廃墟からの復興と現在に至るまでの経済、産業、社会の変遷を、広告用語を通して、現象論的な特徴を明確にすることを目的に研究を行った。

本稿では、広告業のマーケットコミュニケーションを通じての人間社会、環境への適応への貢献の可能性を探るため、企業製品の消費者への宣伝という狭義の広告に止まらず、産業界の発展や、社会情勢を正しく伝えることで、次の10年期への経済・社会動向、企業意識、さらに環境や国際問題等の理解に日本の広告活動が果たしてきた状況について、半定量的なビブリオメトリーで明らかにした結果を報告する。

1. Introduction

Words and languages are the most fundamental preposition that human being could be really “to be human”. We find various speeches and lines such as “Man is social animal” and “Human society consists of relationship among persons”. All these kinds of sentences are closely related to the importance of interactive communication of men and societies with the use of words and languages. However, we have encountered with various mismatches in communication, e.g., in politics, business, and even science and technology. Economic activities in practice are performed and realized holistically by incorporating manufacture, supply and utilization of products, appropriate services in the market, so evaluation, investment and planning for activities where proper information play an important role to heritage these items holistically. The economic activities, however, still tend to be understood mainly from the standpoint of sales and profits, but its mission should be construction of the material base for the happiness of man and society. It could be achieved by keeping the right relationships with human, society and the environment, through promoting the information transparency between producers and consumers.

Taking into consideration the above-mentioned background, the mission of the advertising business in the era of highly developed information technology, leads to the development of the economy and the welfare of society with so-called environment-conscious ways by selecting proper messages of right words and sentences. This paper deals with the historical and phenomenological features of 47 years advertisement activities extracted bibliometrically from only one specialized journal on advertisements as the

first step of our long-term research project.

2. Objective of research

In this research, the whole history of growth and development of advertising business in Japan of 47 years after the end of the World War II has been studied, and important roles of advertising business and the problems left behind have been investigated in relation to the process of economy to recover from the ruins of the war, restart its economic activities with high growth rate driven by the development in heavy industry, and promote mature consumer economy and intellectualization and informatization of society.

For our research, the unique reliable monthly journal, SENDENKAIGI (meaning “Publicity Meeting” in Japanese) has been used as the data source for bibliometric analysis, where particular importance has been paid to such a variety of viewpoints as promotion of business results, mass and consumer economization and the social and environmental relations which are indispensable for the human life in general.

The objects of our study were the words which appeared in the sentences of titles for the contents of the journal during 47 years. Based on these data of words, characteristic phenomena and properties in the advertisement of the post-war Japan have been considered and discussed.

3. Methods of research

As mentioned above, the fundamental data were taken from all the issues of monthly journal of SENDENKAIGI edited from 1954 to 2000.^[1]The journal started to be published, just when the country had

resuscitated from the ruins, by those people who had worked for advertising in private companies, with the perspective that the age of information and communication finally would come to Japan, while reviving and catching up the advanced Western countries. Since then, it has occupied always the position of the most reliable magazine for advertising and publicity in this country, and it is now distributed in 100,000 copies monthly and thought to be indispensable for all the people who work for advertisement and publicity. From the academic point of view, the journal SENDENKAIGI is a huge intellectual mine where there are all kinds of words on advertisement and publicity in Japan, which reflect the trend and characteristics of respective ages.

Firstly, all the titles of the reports and articles were decomposed into individual words and the other elements like punctuations. Then, after removing the meaningless elements of auxiliary words like “Jyoshi”, postposition, conjunctions, adverbs and adjectives, and names of persons and places, 400 frequently used words have been taken up as samples.

On these 400 words, the frequency of appearance has been counted and they are classified into the groups of category based on the applied fields. The age was also classified into the following five decades: 1954/1960, 1961/1970, 1971/1980, 1981/1990, 1991/2000. In each decade, the words of which the frequency were included within the 1st to 100th have been chosen as the most important ones for the advertisement of the decade concerned.

Furthermore, 8 categories have been defined in order to characterize the decades respectively and to investigate the trends of advertisement in relation to the correspond-

ing background of the age. The category consists of the following eight class: (A) media, (B) concepts of advertisement, (C) jobs and business for advertisement, (D) methods and techniques for advertisement, (E) fundamental concepts and activities in advertisement, (F) object of advertisement, (G) business management and activities in general, (H) society and human. (Shown in Table 1.)

4. Results of bibliometric measurement

4.1. Classification of the most frequently appeared words used in the titles of the journal “SENDENKAIGI” from 1954 to 2000.

The result of classification is shown Figure 1 with corresponding each decade. The group of words on the concept of advertisement was 30% of the total, and that of the words on the business management and activities for advertisement was 20~25%. Those two groups were the most used as a single class. It could be understood that the business companies have continued to put much importance on activities toward consumers and societies.

Then, there followed the group of the words on general concepts and activities of business such as research, problem, investigation, and information. They are regarded as important commonly, not only in the sector of advertisement but also in various other business fields.

Particularly, the class H, i.e. “Society and Human,” was at first on a rather lower level because Japan after the war had no room except for obtaining just what people could live their life on. However, with the lapse

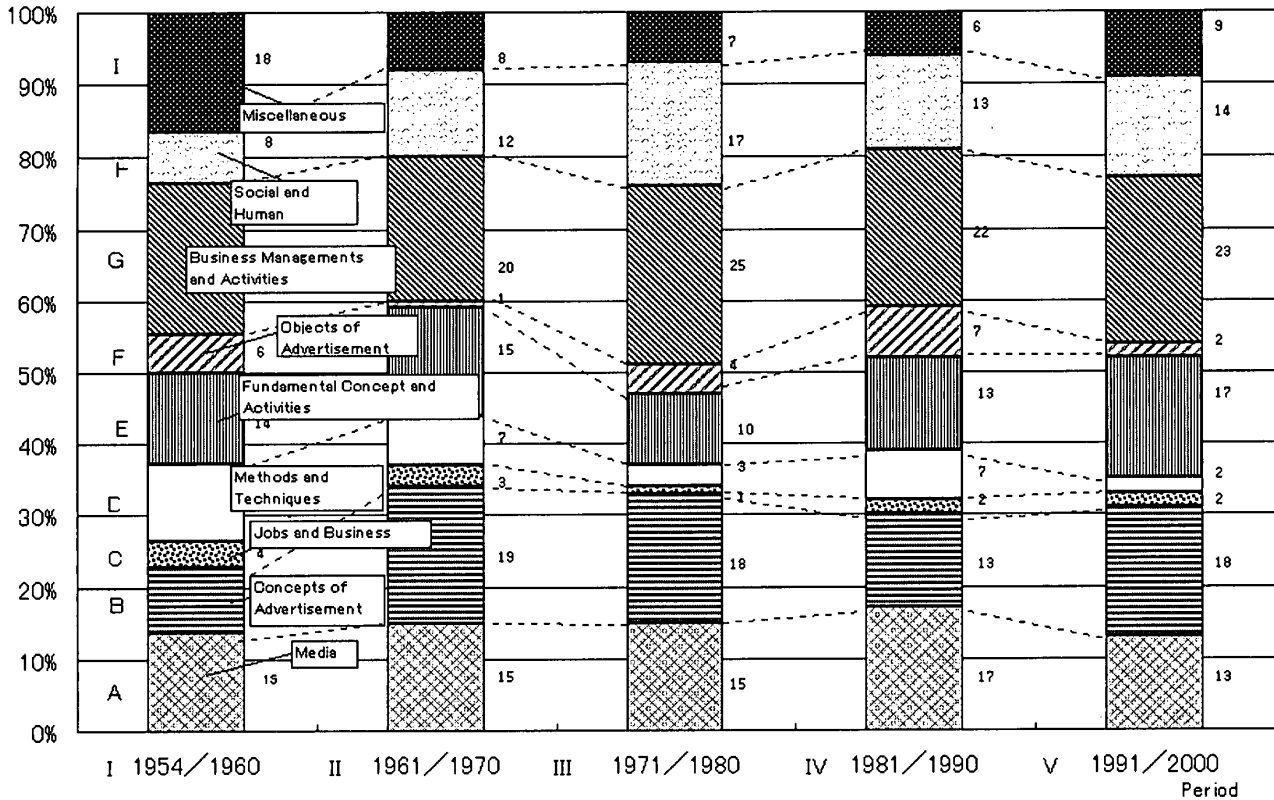


Figure 1: The Summarization of the Results Obtained by Bibliometric Investigation on the Titles of SENDENKAIGI Journal from 1954 to 2000.

of years thereafter, the position of the class H had gone up due to the improvement of their living and reached 15%. These results above are considered to come from that the smooth circulation of information and communication has been regarded as the most important for actual business to create and maintain wide and healthy relationship between suppliers and consumers.

4.2. Chronological change in the usage of letters in the advertisement and after the World War II

Japan had attained the economic recovery under the strong aid and support from the Western countries, especially from the United States. During this period of restora-

tion, the country excluded or even denied many traditional and historical ways of thinking from its cultural policy, in education, public service, social activities and so on. Japan has imported widely various foreign words from Western countries and expressed those words with katakana, the second Japanese phonograms and Roman alphabet itself. One can sometimes hear lamenting voices about the collapse of the original Japanese language system. Especially, such a trend had been obvious in the fields like high technologies such as information and communication and those of fashion, art, and sports for younger generations. In a strong flow of mass- or consumer economy, more and more American and European words have been steadily introduced. Figure 2 chronologically summa-

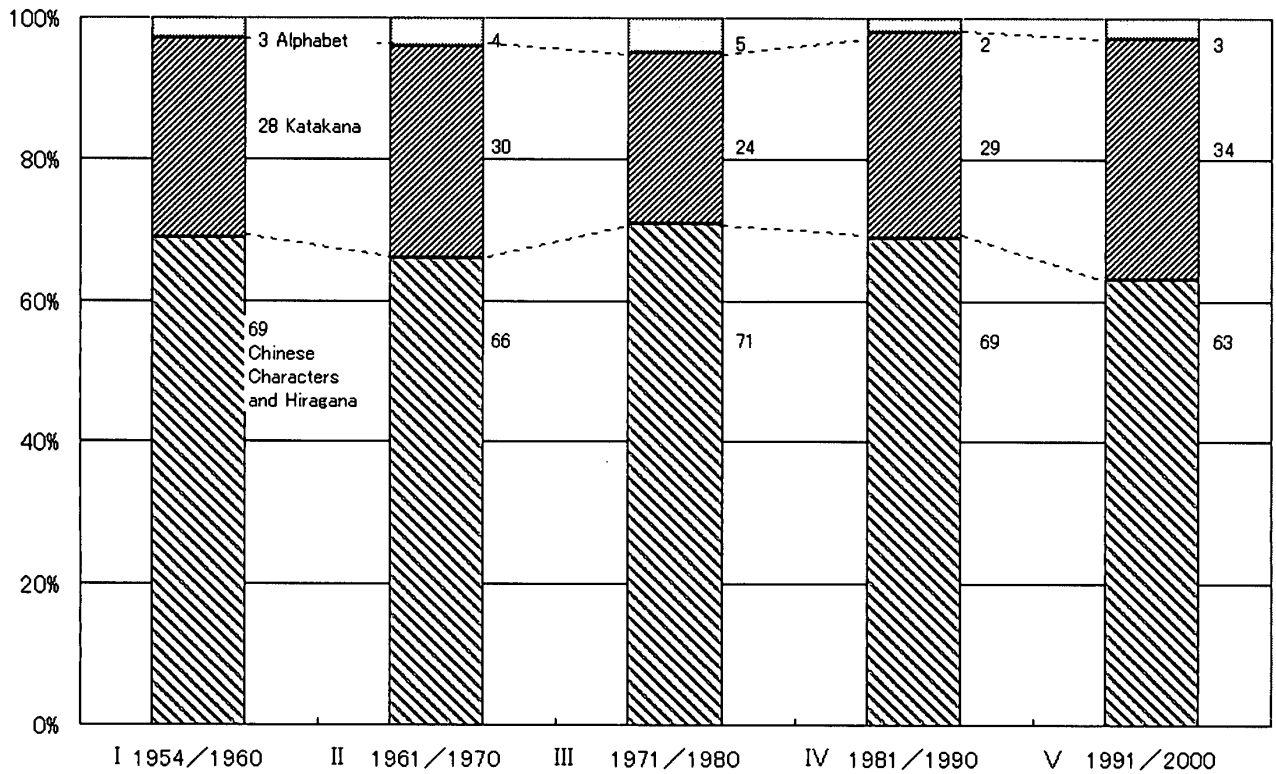


Figure 2: Chronological Change in the Usage of Letters and Characters of the Title Sentences in SENDENKAIGI Journal.

rizes the above-mentioned trend concerning 100 important words of each decade.

As a whole, it is a little surprising and unexpected that the percentage of “Hiragana”, the first Japanese phonogram and “Kanji” Chinese characters, have been almost in the similar level, that is to say, from 65% to 70% up to the decade of 1980s, and in the decade of 1990s it fairly decreased to 63%, which might be thought to correspond to the age of advanced technology, information technology, and the consumer economy and the society. On the other hand, the expression with katakana, another phonogram to express the occidental pronunciation, unexpectedly had kept nearly 30% from the 1950’s to 1980’s, and the direct expression with alphabet had been about 5%. However, in 1990’s, the katakana expression of foreign language has tended to increase and

become 34%, and when the alphabetic expression is included, foreign words have approached the level of 40%. Considering the increasing trend of katakana and alphabet, from 31% at the first decade to 37% at the latest decade, we can clearly confirm that the expression of Japanese in the advertisement and publicity has steadily and clearly changed after the World War II.

Therefore, corresponding to the developing informatization and globalization of the current age, the traditional Japanese expression will decrease, down below the level of 50% in the near future. It is anticipated that such a result might trigger the bigger change in the general structure of the Japanese language, not only in the advertisement.

4.3. Chronological change of the words used in the advertisement of each category

4.3.1. Change in Japanese words for advertising media

In Table 1, which summarizes the result of the bibliometric analysis about the words on the advertising media, it is unexpectedly noted that, although information media are said to have been drastically changed by technologies of electric wave and electronics, such traditional media as newspaper, book, journal, i.e., paper media, have situated themselves in the top group. Certainly electric wave and electronics have been effective and useful for multimediatization with its variety of appeals to consumers' sense of hearing and seeing, but on the other hand, the traditional printed media have occupied the upper position because of their repeatability of appeal and their persuasiveness, both of which enable consumers and citizens to deepen their understanding on their consuming products.^{[2][3]} Furthermore, such information technology as electronics, computer, and networking, and new media like television, video, network, CATV, were introduced into the advertisement from the decade of 1980's, and particularly in 1990's when Internet, the leading technological element of the networked society, has begun to be widely used in the advertising business.

In addition to this, it is also noteworthy that "everydayness" of newspaper and weekliness or monthliness of journal, i.e., the periodicity with repeatability and reproducibility brings the strong potentiality to the media for advertisement.

4.3.2. Changes in Japanese words for advertisement on specific concepts

As seen in Table 2, "Advertising" have occupied the top position through the whole period, from 1954 to 2000, and "Propaganda" and "CM" were always positioned next to "Advertising" because SENDENKAIGI is the unique journal of advertisement specialty and the concepts related to this specialty have been always the most important. Further, "Copy" and "PR" which have now become daily words, appeared first at the upper position at the decades of 1950's, as a result of importing the American publicity and consumerism culture. It is also to be noted that although the number of the words of this class were only 10 in the decade of 1950's, not so long after the end of the war, it had already increased twice in the next decade because of the rapid recovery of Japan's economy.

The words which are now widely seen and heard in the advertisement and publicity like "Campaign," "Image", "Illustration", "Communication" and "Creative", had been already used much frequently in 1960's, and it also comes from the positive introduction of American culture of advertising business.

4.3.3. Change in Japanese words for advertisement and publicity related on specific jobs and business, and methods and technologies

At the first stage of Japan's advertising industry, most of works had been pursued by the people in the advertising section of manufacturing or sales companies, or by the staff of advertising companies. In the decade of 1950's, there had already appeared specialists of advertisement. In this decade, as

Table 1: Chronological Change from 1954 to 2000 in the Usage Frequency of the Words in the Titles of SENDENKAIGI Journal: Media.

Period Rank	I(1954/1960)	II(1961/1970)	III(1971/1980)	IV(1981/1990)	V(1991/2000)	VI Total
1	Television	Books	Books	Media (in English)	Books	Books
2	Newspaper	Newspaper	Newspaper	Books	Media (in English)	Media (in English)
3	Radio	Television	Television	Broadcasting	Television	Television
4	Books	Magazine	Media (in English)	Newspaper	Newspaper	Newspaper
5	Programs	Radio	Radio	Television	Broadcasting	Broadcasting
6	Magazine	Media	Novel	TV	Magazine	Magazine
7	TV	SENDENKAIGI	Magazine	Communications	SENDENKAIGI	SENDENKAIGI
8	Media	Broadcasting	SENDENKAIGI	Magazine	Media	Media
9	Media (in English)	Media (in English)	Private Broadcasting	Publishing	Mass Communication	Mass Communication
10	SENDENKAIGI	Mail	Media	Radio	Communications	Communications
11	Broadcasting	Novel	Publishing	Media	Internet	Internet
12	Asahi	Movie	Communications	Videa	TV	TV
13	Movie	Asahi	Video	Press	multi-media	multi-media
14	Poster	Direct Mail	Movie	Net		
15	Private Broadcasting	Poster	DM	SENDENKAIGI		
16				Network		
17				CATV		
Total	15	15	15	17	13	13

shown in Table 3, appeared such words as “Copywriter”, “Director”, and “Designer” and they are very widely used for advertising business in Japan even today. In the following decade of 1960’s, “Adman”, and “Character” had been already used frequently. These explain specialization and independence of jobs and business in Japan’s advertising field.

Table 4 gives the changes in the words re-

lated to methods and technologies for advertisement. Most of these had been imported in the first decade, during which the free market economy had begun after the end of the war. As those words had gotten so rapidly widespread in practice and become too much popular in a sense, the frequency of their appearance in SENDENKAIGI on the contrary had tended to gradually decrease because the journal aims to provide discus-

Table 2: Chronological Change from 1954 to 2000 in the Usage Frequency of the Words in the Titles of SENDENKAIGI Journal: Concepts of Advertisement.

Period Rank	I(1954/1960)	II(1961/1970)	III(1971/1980)	IV(1981/1990)	V(1991/2000)	VI Total
1	Advertising	Advertising	Advertising	Advertising	Advertising	Advertising
2	Propaganda	CM	Campaign	CM	CM	CM
3	CM	Propaganda	CM	Propaganda	Professional	Propaganda
4	Copy	Copy	Propaganda	Copy	Communication	Professional
5	Professional	Campaign	Copy	Professional	Propaganda	Copy
6	Image	PR	Professional	Communication	Campaign	Campaign
7	PR	Image	Creative	Event	Research	Communication
8	Commercial	Professional	Monitoring	Campaign	SP	Net
9	Awards	Illustration	PR	NEWS	Event	PR
10	Production	Communication	Communication	SP	Brand	Event
11		Creative	Production	Monitoring	Production	SP
12		Art	Program	Image	PR	Production
13		News	Image	PR	Promotion	Research
14		Program	Event		Copy	Creative
15		Awards	NEWS		Adweek	Promotion
16		Editing	Advocacy Ad		Program	Brand
17		Commercial	Propaganda (in English)		Printing	News
18		Production	Promotion		Image	
19		Festival				
Total	10	19	18	13	18	17

Table 3: Chronological Change from 1954 to 2000 in the Usage Frequency of the Words in the Titles of SENDENKAIGI Journal: Jobs and Business for Advertisement.

Period Rank	I(1954/1960)	II(1961/1970)	III(1971/1980)	IV(1981/1990)	V(1991/2000)	VI Total
1	Copywriter	Ad man	Copywriter	Copywriter	Copywriter	Copywriter
2	Director	Copywriter		Character	Character	
3	Designer	Designer				
4	Tie-up					
Total	4	3	1	2	2	1

sion by the specialists.

On the other hand, it is very interesting too, that from the 1980's, words like "Music" and "Image", which imply the age of multimedia, have been very popular.

4.3.4. Changes in Japanese words frequently seen on fundamental concepts and business activities of advertising

The advertising business, as all the other business, aims to contribute to the wealth

Table 4: Chronological Change from 1954 to 2000 in the Usage Frequency of the Words in the Titles of SENDENKAIGI Journal: Methods and Techniques for Advertisement.

Period Rank	I(1954/1960)	II(1961/1970)	III(1971/1980)	VI(1981/1990)	V(1991/2000)	VI Total
1	Design	Design	Work	Design	Idea	Design
2	Evaluation	Expression	Expression	Image	Expression	Expression
3	Words	Works	DM	Naming		Idea
4	Art	POP		Expression		
5	Works	Dictionary		Music (in English)		
6	Outdoor Advertisement	DM		Corporate Identity		
7	Lettering	Idea		Music		
8	Photo					
9	Idea					
10	Graphic					
11	Typografi					
12	Catch Phrase					
Total	12	7	3	7	2	3

and happiness of people and society in the harmonious correlation between business and consumers. Particularly, advertising business must pay much effort to take the appropriate way of conduct from the standpoint of harmonic and holistic consideration.

Therefore, accommodation and adaptation of the advertising business to the age and the environment is always the indispensable theme of the journal to be dealt with, particularly on the basis of fundamental concepts, on technologies and various social relationships. As shown in Table 5, in the upper group during the whole period were found "Information", "Research", "Study", "Problem", and "Training." All these words are thought to have been used so much because they are always related to all kinds of business activities in common.

In spite of many voices recently heard, the number of IT-related words was unexpect-

edly small in the decade of 1990's. They are only, "Digital", "Computer" and "Personal Computer." It implies that the technology and products expressed by these words have been already too much widespread and became so well known and so that they have lost the appealing capability for advertising, that is to say, they have become simple proper nouns.

4.3.5. Changes in Japanese words on advertising objects

The journal SENDENKAIGI pays much importance to the themes of field research and practical discussions of common products in the market to be targeted in advertisement, but mostly it is not a journal focusing on the individual advertisements of concrete products. The names of products that appeared in this bibliometric study, therefore, are the names of principal and common

Table 5: Chronological Change from 1954 to 2000 in the Usage Frequency of the Words in the Titles of SENDENKAIGI Journal: Fundamental Concepts and Activities in Advertisemet.

Period Rank	I(1954/1960)	II(1961/1970)	III(1971/1980)	VI(1981/1990)	V(1991/2000)	VI Total
1	Study	Problem	Information	Information	Information	Information
2	Lecture	Study	Problem	Study	Data	Study
3	Problem	Research	Lecture	Training	Digital	Lecture
4	Research	Lecture	Study	Lecture	Effect	Problem
5	Science	Technique	Research	Problem	Lecture	Research
6	Training	Activity	Task	Research	Research	Data
7	Promotion	Effect	Training	Idea	Movement	Effect
8	Technique	Data	Activity	Task	Method	Activity
9	Task	Training	Movement	Science	Activity	Task
10	Vision	Task	Creation	Movement	Task	Movement
11	Research (in English)	Creation		Technique	Life	Digital
12	Information	Analysis		Hit	Study	Technique
13	Subject	Information		Data	Training	Analysis
14	Education	Change		Effect	Analysis	Creation
15		Science			Hit	
16					Computer	
17					Task	
18					Personal Computer	
Total	14	15	10	14	18	14

products, considering the role and meaning in advertisement.

The Table 6 shows that there were two characteristic periods in terms of advertisement on products. The first one is the decade of 1950's, the period of recovery after the war, and the second one is the decade of 1980's. In the latter, the Japan had caught up with the economical top group of countries in the world, and the consumer economy had widely developed, corresponding with the high growth of Japanese economy. In the former period, names of daily goods

for living such as drug, car, liquor, and cosmetics appeared frequently. But in the other period, words like "Sports" and "Game" related to leisure and enjoyment of life and words like "Beer" and "Food" related to table luxuries appeared frequently, reflecting the improvement of living standard and the affluence of the society in the country.

Besides, from the continual appearance at the upper position of the words related to the automobile, we can recognize anew that the latter half of the 20th century was just the "century of car" for Japan, too.

Table 6: Chronological Change from 1954 to 2000 in the Usage Frequency of the Words in the Titles of SENDENKAIGI Journal: Objects of Advertisement.

Period Rank	I(1954/1960)	II(1961/1970)	III(1971/1980)	VI(1981/1990)	V(1991/2000)	VI Total
1	Drug	Car	Car	Sports	Car	Car
2	Car		Automobile	Car	Transportation	Sports
3	Liquor		Cosmetics	Game		
4	Cosmetics		Sports	Beer		
5	Automobile			Suntory		
6	Transportation			Food		
7				Automobile		
Total	6	1	4	7	2	2

4.3.6. Changes in Japanese words on business management and activities

As the advertisement is understood to be an interfacial communication, in which the business side should always consider the society, business activities in management, have been often dealt with in SENDENKAIGI from various standpoints and purposes. As seen in Table 7, it could be well understood that in the top six, there exist “Market”, “Marketing”, “Company”, “Strategy”, “Corporation”, and “Merchandise”. It could be said from the view point of management, what is most important is “How a company or corporation sells or does marketing in the market their products, and what kind of strategy should be applied.”

The words that appeared in every decade were “Meeting”, “Consumption”, “Planning”, “Company”, “Marketing” or “Market”, and “Corporation”. All of them are related to the most important themes that are required to find practical solutions corresponding to the environment and the period concerned.

4.3.7. Change in Japanese words related to society and human

In the last category, there are notable

words related to society and human. The advertisement drives the circulation of information in the concerned economy system and there have been innumerable cases of misuse of advertisement, which have led the citizen or mass to the evil trend of society. In the journal of SENDENKAIGI, various researches and investigations have been performed, on the influence of advertisement on society, human, and relationship with the age and the environment.

We find in Table 8, “Times”, “Abroad”, “World”, “Society”, and “Spirit” at the upper position, which reflects the consistent policy of advertising companies to pursue their activity toward the world, while getting rid of the traditional insularism of Japan. These words have been always important key words in every decade and are increasing more and more importance now and in the coming age.

Finally, even in the world of advertisement, strong interest in the influence on society and human have been clearly confirmed, in terms of the frequency of appearance in SENDENKAIGI of such words as “Culture”, “Women”, “City”, “Future”, and “Environment”. It comes from the social trend from the high economic growth, then consumer society and up to information society of today.

Table 7: Chronological Change from 1954 to 2000 in the Usage Frequency of the Words in the Titles of SENDENKAIGI Journal: General Business Management and Activities.

Period Rank	I(1954/1960)	II(1961/1970)	III(1971/1980)	VI(1981/1990)	V(1991/2000)	VI Total
1	Sale	Strategy	Marketing	Company	Marketing	Marketing
2	Management	Meeting	Company	Strategy	Company	Company
3	Sales (in English)	Company	Corporation	Marketing	Corporation	Strategy
4	Agency	Merchandise	Strategy	Consumption	Strategy	Corporation
5	Meeting	Market	Meeting	Development	Trade	Merchandise
6	Consumption	Consumption	Market	Products	Market	Market
7	Planning	Agency	Consumption	Distribution	Meeting	Meeting
8	Corporation	Dentsu	Business	Market	Business	Consumption
9	Market	Industry	Agency	Business	Merchandise	Trade
10	Activity	Sale	Development	Corporation	Consumption	Business
11	Merchandise	Products	Dentsu	Trade	Dentsu	Dentsu
12	Debtsu	Sales (in English)	Planning	Industry	Planning	Sale
13	Organization	Plan	Merchandise	Meeting	Activity	Planning
14	President	Tactics	Sale	Sale	Organization	Development
15	Company	Handling	Handling	Management	Personnel	Managemnet
16	Manufacturing	Planning	Trade	Planning	Businessman	Industry
17	Plan	Market (in English)	Hakuhodo	Scene	Management	Products
18	Development	Management	Industry	Enterprise	Development	Agency
19	Industry	Trade	Distribution	Shiseido	President	Distribution
20	Sponsor	Corporation	Tactics	Dentsu	Hakuhodo	Sales (in English)
21	Tactics		Managemnet	Sales	Sponsor	Hakuhodo
22	Consumer (in English)		Product	Manufacturing	Industry	Organization
23			Scene		Top	
24			Sales (in English)			
25			Manufacturing			
Total	23	20	25	22	23	22

Hereafter, more and more new words with holistic meaning are expected to appear in advertisement, directing toward the harmo-

nious development of economy with the sufficient consideration paid to the importance of society and human.

Table 8: Chronological Change from 1954 to 2000 in the Usage Frequency of the Words in the Titles of SENDENKAIGI Journal: Society and Human.

Period Rank	I(1954/1960)	II(1961/1970)	III(1971/1980)	VI(1981/1990)	V(1991/2000)	VI Total
1	Spirit	Spirit	Times	Times	Times	Times
2	World	Times	Abroad	Abroad	World	Abroad
3	Psychology	Abroad	Society	Life	Society	World
4	Times	World	Today	Circumstances	Movement	Society
5	Today	Psychology	World	Society	Circumstances	Spirit
6	Love	Environment	Environment	Region	Environment	Circumstances
7	Abroad	Society	Rules	City	Popularity	Life
8	Society	Life	Spirit	Culture	Reform	Environment
9		Future	Local	Movement	Regulations	City
10		Student	Culture	Women	Spirit	Regulations
11		Circumstances	Circumstances	Spirit	City	Culture
12		General Public	Expression	World	Future	Women
13			General Public	War	Love	Subject
14			Human		Women	Future
15			Life			
16			Nazi			
17			War			
Total	8	12	17	13	14	14

5. Discussions

First, to follow the appearance and disappearance of the words for advertisement business, the words which had been included in the group of the upper 100 but disappeared in the following decades, were listed up. They had been used and subjected to research and discussion in the journal for a certain period of time, but, after having played their roles, they had dropped out of the top 100th position. On the contrary, new important themes have emerged and they brought in newer expressions by introducing new words.

As seen in the tables from 9 to 13, the columns marked with sign (+) gives the number of words newly appeared in the concerned decade, and the columns marked with sign (-) the number of words which disappeared from the same period and went out of 100th position in turn. It corresponds to the rise and fall of the frequency in the usage of advertising-related words. Particularly, the words in both columns of (+) and (-) at the same decade imply how the people in the advertising field so strongly paid much importance to the characteristic themes and timely problems concerned.

Table 9: The Rise and Fall of Usage Frequency of Words in the Titles of SENDENKAIGI Journal: Media.

Period Rank	I(1954/1960)		II(1961/1970)		III(1971/1980)		IV(1981/1990)		V(1991/2000)	
	-	+	-	+	-	+	-	+	-	+
1	TV	*Mail	Mail	*Private Broadcasting	Novels	•TV	Publishing	Mass Communication		
2	Poster	Novels	Asahi	Publishing	Private Broadcasting	Press	Radio	Internets		
3	Private Broadcasting	*Direct Mail	Direct Mail	Communications	Sales	*Net	Video	Multimedia		
4			Poster	Video		*Network	Press			
5						CATV	Net			
6							Network			
Total	3	3	4	4	3	5	6	3		

Table 10: The Rise and Fall of Usage Frequency of Words in the Titles of SENDENKAIGI Journal: Concepts of Advertisement.

Period Rank	I(1954/1960)		II(1961/1970)		III(1971/1980)		IV(1981/1990)		V(1991/2000)	
1	-	+	-	+	-	+	-	+	-	+
2		*Illustration	Campaign	Monitoring	Creative	Event	Program	News (in English)	Monitoring	Research
3		Communi- cation (in English)	Illustration			Advocacy Ad				Brand
4		Creative	News	News (in English)	Advocacy Ad	Propaganda (in English)				Production
5		Art	Editing	Advocacy Ad	Propaganda Promotion					•Promotion
6		*Editing	Festival	*Promotion	Production					ADWEEK
7		*Festival								•Program
		Program								Printing
Total	0	7	5	6	6	6	0	2	7	

Table 11: The Rise and Fall of Usage Frequency of Words in the Titles of SENDENKAIGI Journal: Fundamental Concepts and Activities of Advertisement.

Period Rank	I(1954/1960)		II(1961/1970)		III(1971/1980)		IV(1981/1990)		V(1991/2000)	
		-	+	-	+	-	+	-	+	-
1	Vision	Activity	Technology	*Movement	Activity	Idea	Problem	Digital		
2	Research	*Effect	Effect		Movement	*Science	Idea	Method		
3	Topic	*Data	Data		Creation	*Technology	Science	Activity		
4	Education	Creation	Analysis			Hit	Technology	Topic		
5		*Analysis	Change			Data		Life		
6		*Change	Science			Effect		●Analysis		
7								Computer		
8								Personal Computer		
Total	4	6	6	1	3	6	4	8		

Table 12: The Rise and Fall of Usage Frequency of Words in the Titles of SENDENKAIGI Journal: General Business Management and Activities.

Period Rank	I(1954/1960)		II(1961/1970)		III(1971/1980)		IV(1981/1990)		V(1991/2000)	
	-	+	-	+	-	+	-	+	-	+
1	Shift	Strategy	Merchandise		Agency	Sale	Products		●Marchandise	
2	Organization	Products	Plan	Development	Dentsu	Enterprise	Distribution	Activity		
3	President	*Market	Market	*Hakuhodo	Sale	Shiseido	Sale	Organization		
4	Manufacturing	the Trade		Distribution	Handling	*Sales	Scene	Personnel		
5	Development			Scene	Hakuhodo	*Manufacturing	Enterprise	Businessman		
6	Sponsor			Manufacturing	Tactics		Shiseido	President		
7	Consumer				Sales (in English)		Sales	●Hakuhodo		
8	Marketing						Manufacturing	Sponsor		
9								Top		
Total	8	4	3	6	7	5	8	9		

Table 13: The Rise and Fall of Usage Frequency of Words in the Titles of SENDENKAIGI Journal: Society and Human.

Period Rank	I(1954/1960)		II(1961/1970)		III(1971/1980)		IV(1981/1990)		V(1991/2000)	
		-	+	-	+	-	+	-	+	-
1	Today	Environment	Psychology	Environment	Today	*Rules	Abroad	*Region	●Environment	
2	Love	Life	Future	Life	Environment	*Local	Life	City	Popularity	
3		*Future	Student	*Future	Rules	Culture	Region		Reform	
4		*Student		*Student	Local	*Regulations	Culture		●Regulations	
5		Circumstances		Circumstances	Regulations	*Human	Spirit		●Future	
6		General Public		General Public	General Public	Nazi	War		●Love	
7					Human	War				
8					Nazi	*Today				
Total	2	6	3	8	8	2	6	6		

Table 14: Summarization of the Rise and Fall of Usage Frequency of Words in the Titles of SENDENKAIGI Journal.

Category	Period	I	→	II	→	III	→	IV	→	V				
		1954/1960		1961/1970		1971/1980		1981/1990		1991/2000				
1. Media		(6)	—	7	—	(8)	—	8	—	(9)	—	11	—	(9)
2. Concepts of Advertisement		(7)	—	13	—	(12)	—	12	—	(6)	—	12	—	(9)
3. Fundamental Concepts and Activities of Advertisement		(10)	—	12	—	(7)	—	4	—	(7)	—	10	—	(12)
4. General Business Management and Activities		(12)	—	7	—	(9)	—	13	—	(12)	—	13	—	(17)
5. Society and Human		(8)	—	9	—	(11)	—	16	—	(10)	—	9	—	(9)

First, on the results in Table 9 on advertising media, “Television” had already been appeared so early and it had been rapidly and widely introduced into the Japanese society.^[4] Television has naturally become as one of the most media for advertisement during the first decade, when private broadcast systems started the commercial-base broadcasting, the business budget of which had been based on airing various advertisements. That is why “Private broadcasting” appeared frequently from 1950’s to 1960’s, which was the beginning period of the growth of Japan’s economy after the end of the war, with many American ways and techniques for advertisement introduced.

By observing the turnovers of words in the respective decades, as a whole, the changes of advertising media with development of the technology and economy of Japan could be clearly understood. For example, innovation in information and communication technology had led to the introduction of words such as “Video”, “Net”, “CATV”, “Network” and “Internet”, and the words like “Press”, “Private Broadcasting” had finished the main role till then. However, it did not mean that these words could not be seen in the practices of advertising business at all. However, as SENDENKAIGI was the journal

for advertising specialists, the words that had disappeared were replaced by the other next-coming words to deal with more up-to-date themes on advertisement. Therefore, it should be understood that such a medium as car radio has recently come to be widely used again for advertisement.

Furthermore, the words marked with asterisk (*) were the words newly entered and used much frequently during the concerned decade, but they had gone out of the top 100 words group in the following decade. That is to say, these words had been regarded as so important and used so frequently for advertisement, and consequently, their concepts or techniques had attained a level to be very common in daily life. For example, “Mail” and “Direct mail” had been appeared and competed with the general post-mail, and a so-called boom of direct mail had come and expanded in the daily life. “Private broadcasting” in the 3rd decade of the high economic growth, “Net” and “Network” in the 4th decade have been considered to come from the similar situation. And (●) mark on “TV” denotes that the word reappeared in the top 100 word group with a new meaning or role for the advertisement.^[5]

This similar trend could be observed in Table 10 on concepts of advertisement, and in

Table 11 on fundamental concepts and activities of advertisement, in Table 12 on general business management and activities, and in Table 13 on society and human.

For example, the number of rise and fall of the words observed in Table 12 means the characteristic of the decade concerned because of its direct relationship with business activities. Particularly with collapse of the so-called bubble economy and under the new wave toward the big competition and the globalization in the world economy, there had been appeared in the 5th decade new words to express the important elements and conditions for the recovery and restructuring of Japanese economy, such as “Merchandise”, “Activity”, “Organization”, “Personnel” and “Businessman”. As known from Table 13, the interest of advertisement in society and human had changed with the lapse of year. The most interest in the decade of 1960's, when the economy had fundamentally entered the recovering stage, was business growth, mass-economization and social phenomena like student movement. Then in the following decade of the high economic growth, such words as “Culture” and “Human” become the point of focus because of the influence of materialistic affluence of the society. The words like “Nazi” and “War” are thought to have appeared because of the cold war between two superpowers in the world at that time.

As regards the environment, so-called pollution or public nuisance were important issues even for a journal about advertisement like SENDENKAIGI, and as the result of the application of positive environmental policies and technologies, words related to the environmental pollution in this sense seemed to have finished their main roles in this field of advertisement. However, “Environment”

has reappeared in the top 100 words group during the decade of 1990's, as the global environmental issues became so severe for mankind and nature. Besides such a background, “Future” and “Love” have also reappeared because the period has a very important meaning from the viewpoint of the transition from 20th to 21st century.

Table 14 summarizes the in-and-out of the words in the tables from 9 to 13. The figures in parenthesis denote the total number of the words that went out in each decade and those that newly appeared. This figure could be regarded as a kind of measure to judge the dynamic situation of the advertising business field. To summarize, the turnovers of words for advertisement were very active in various fields such as media, concept of advertisement, its fundamental concept and activities, society and human, and so on. On the other hand, viewed along the lapse of time, the Japanese advertising activities had begun to increase its dynamism remarkably from the decade of 1970's.

6. Conclusion

The frequency of the word usage in advertisement has been studied by means of bibliometric analysis, looking into the titles of SENDENKAIGI, from 1954 to 2000, the unique journal for advertising specialty in Japan. As a result, the trend and change of the advertising activities in Japan after the World War II have been made clear both quantitatively and schematically. The main results are concluded as follows;

- 1) The characteristics of advertisement in Japan have schematically and quantitatively corresponded to the economical recovery and development of the concerned age in respective specific fields.

- 2) The trend of westernization through infiltration of occidental words into the Japanese advertisement after the World War II, has unexpectedly not so drastically proceeded as generally supposed. However, the traditional Japanese expression have gradually decreased and approached just the level of 60% at the end of the 20th century. If it reached under 50%, a big change in the structure of Japanese language might be triggered.
- 3) The words much frequently used (within 100) for advertising in Japan after the World War II have been studied concerning 8 groups, where the post-war recovery and growth in Japanese economy have been well reflected in the frequency of the their usage.
- 4) The trend of appearance and disappearance of words of each decade has made clear semi-quantitatively a dynamic relationship between the advertisement and the economic situations of the related decades in Japan.

It could be said that from the bibliometric measurement of frequency of words which have been appeared in the table of contents in SENDENKAIGI, the only one journal of advertisement specialists, has made us understand phenomenologically the changes of themes and backgrounds which have been important for the advertising field in Japan after the World War II.

In the following step, the cluster analysis of the words are now being studied to make clear directing factors and their relation in order to grasp the coming direction of business activity in Japanese economy.

Acknowledgement

At the end of our thesis, I would like to

thank all those people who made this thesis possible and an insightful experience for us.

Especially we wish to express our sincere gratitude to Nobuto Oka, Hayato Amioka and Kohei Yamada, who conducted the bibliometric research on the journal, looking through the entire 47-year-history of the publication.

References

- [1] Monthly SENDENKAIGI, Vol. 1 (March 1954)–Vol. 613 (December 2000)
- [2] Tasaburo Kobayashi, “Kokoku to Kankyo: Kokoku Kino no Jisshoteki Bunseki”, Jikkyo Shuppan, Tokyo (1975)
- [3] Tasaburo Kobayashi et al, “Kokoku Communication Ron”, Zeimukeiri Kyokai, Tokyo (1977), p1–23
- [4] Marshall McLuhan, “Understanding Media”, McGraw-hill Book Company, New York (1964), [Trans. Misuzu Shobo, Tokyo (1987)]
- [5] Yasuhiko Kobayashi, “Kokoku: Mou-Hitotsu no Kagaku”, Jikkyo Shuppan, Tokyo (1982), p205–223

(2002年5月31日受付)

(2002年7月3日採録)

著者紹介

東 英弥 (あずま ひでや、正会員)

日本大学法学部卒業

東京国際大学 商学研究科 修士課程終了

株式会社 宣伝会議 社長

株式会社 富士テック 会長

国内における代表的な広告宣伝分野の専門誌「宣伝会議」、デザイナー部門の専門誌「ブレン」等を介して、マーケティング、デザイン

界のレベルアップを目指し、また富士テックが担当する環境関連の設備機械の普及拡大により、環境事業を内部経済化することに強い関心を持つ。IT時代における環境分野をはじめ、新しい事業創出と発展に向けて実践的理論研究を行っている。

E-mail: hideazuma@sendenkaigi.co.jp

佐藤 純一 (さとう じゅんいち、正会員)

東京大学工学部冶金学科卒、工学博士
同大学工学部大学院博士課程を経て、昭和電工株式会社に就職し、アルミニウム製錬、ヒートパイプ熱交換器、ファインセラミックス開発・製造を担当し、1993年に退職。1992年から1998年まで、東京大学工学部寄付講座 相関製鉄システム講座の専任客員教授を務める。現在も同大学人工物工学センター客員研究員。1997年より現在技術と人間を哲学するメタテクニカフォーラムを主宰。1997年より国際メタテクニカテクノロジー研究センター所長を務め現在に至る。現在は、メタテクニカの視点から、技術と人間、社会についての研究に力を入れている。

E-mail: H8G01160@nifty.ne.jp

岩田 修一 (いわた しゅういち、正会員)

東京大学教授、工学博士
東京大学工学部原子力工学部卒、同大学博士課程終了

1997年～2000年 東京大学人工物工学研究センター長

現在 同研究センター ライフサイクル工学研究部門教授

合金設計、材料設計、設計支援システム、設計化学、人工物工学、情報学等、伝統的な学問分野として認知されていない分野における求解のプロセスの体系化とシステム開発を行ってきた。成果の一部を、「物質・材料のための仮想実験システム」、「材料メタデータ」、「複雑システムのモデリング技術」、「LPF (Linus Pauling File)」として提示した。平成12年10月より、ICSU 科学技術データ委員会 (CODAT) の執行委員となり、DATA&SOCIETY プロジェクトとして、「社会における意思決定」を適正なものとするための国際的活動を準備中である。

E-mail: iwata@race.u-tokyo.ac.jp

学術雑誌電子化支援システムとしての J-STAGE SGML 編集ツール

The J-STAGE SGML Editing Tool as an Electronization Support System of Academic Journals

中川 理* 西脇 二一†

Osamu NAKAGAWA* and Niichi NISHIWAKI†

* 科学技術振興事業団 情報加工分析部 電子ジャーナル部門

Electronic Journal Division, Department of Processing and Analyzing Information,
Japan Science and Technology Corporation

E-mail: o2nakaga@jst.go.jp

† 奈良大学 社会学部 人間関係学科

Department of Human Relations, Faculty of Social Research, Nara University
niichi@daibutsu.nara-u.ac.jp

学術雑誌の電子ジャーナル化を推進するためには電子化に伴う編集委員会の負担を軽減することが必要である。電子化支援システムの一つとして開発された J-STAGE SGML 編集ツールは、雑誌の印刷後に組み版ソフトから書き出された論文毎のテキストファイルに SGML タグ付けを行うもので、これにより従来の編集・印刷プロセスを変更することなく、J-STAGE 搭載に必要な SGML データを容易に作成することができる。本論では、その構造と機能を紹介し、学術雑誌の電子化の将来について議論した。

It is necessary for the promotion of electronization of academic journals to reduce the load of editorial committee. The J-STAGE SGML Editing Tool was developed as an electronization support system, which adds SGML tags on the text file of each paper that was converted from the DTP software after publication of the journal. It makes possible without any modification of the current editing and printing process to create the SGML data which the J-STAGE requires to accept the journal. In this paper its structure and functions are introduced, together with the discussion on the future of electronization of academic journals.

はじめに

インターネットの普及に伴い、学術雑誌を電子ジャーナル化して公開する動きが広がっている。公開の形式としては、学会独自のホームページで学会誌の内容を公開するもの、印刷会社が担当している学術誌をデータベース化して公開するもの、および、公的機関による電子図書館システムの中で公開するものなどがある。

学会誌の電子ジャーナル化に対応できる印刷会社は限られており、そのための経費も小さくない。規模の大きな学会は印刷会社に冊子体の発注と一緒に委託することもできるが、大多数の学会では学会の中で対応せざるを得ない。

電子ジャーナル化といっても、印刷物から画像ファイルに変換したものを表示・印刷させるだけであれば、電子化による作業量の増

加は少ない。しかし、これでは電子ジャーナル化のメリットである検索や他システムとのリンクなどの機能が使えない。論文を電子化するためには、書誌情報、要旨、キーワード、本文、文献情報などを構造化する必要があり、そのためには相当の作業が必要である。

これらの電子ジャーナル化の作業は編集委員会の担当になるが、多くの学会では編集委員会はボランティア的に運営されており、原稿の募集、受付、査読、審査、編集、校正などの従来の業務で手一杯である。編集委員会の業務に電子ジャーナル化を追加するのであれば、電子化のための支援システムによる作業の効率化が必要である。

本論では、これらの課題を解決することを目的として開発した電子化のための支援ツールである「J-STAGE SGML 編集ツール」を紹介する。

J-STAGE

科学技術振興事業団は、学術雑誌の電子化、科学技術情報の発信・流通の促進をするために、科学技術情報発信・流通総合システム(以下、J-STAGE)を構築した(吉田ほか, 1999; 吉田・尾身, 2001; 時実, 2000; 白木澤, 2000)。2002年6月末において、74誌が公開されている。

J-STAGEでは、各ジャーナル化毎に論文データを記録するためのデータベースを持っている。一つの論文データがデータベースの1レコードに対応する。書誌情報、引用文献に関する情報、全文テキストデータ、および、論文本体の閲覧用のファイルであるPDFが、それぞれのレコードのフィールドに格納されている。これらの論文データを閲覧する際には、その論文の情報がデータベースより動的なHTMLファイルとして生成されることになる。

電子化支援ツール

J-STAGEでは、論文を電子化するための形式の一つとしてSGMLを採用し、多様な学術論文をSGML化するためにJ-STAGE DTDを定義している(科学技術振興事業団, 2002a)。J-STAGEに必要なデータを作成するためには、Adobe FrameMaker+SGMLの汎用テンプレート、Microsoft Access対応書誌事項作成支援ツール、および、J-STAGE SGML編集ツールを用いることができる。

Adobe FrameMaker+SGML 汎用テンプレートはJ-STAGEの標準システムの一部で、版下の作成、PDFの作成、書誌情報用のSGMLデータの作成を同時に行うことができる。しかしながら、版下データの作成方法をAdobe FrameMakerに変更しなくてはならないため、この方法の採用が難しい学会もあった。

Microsoft Access対応書誌事項作成支援ツール(科学技術振興事業団, 2002b)は、それぞれの学会誌の版下作成方法を変更することなく電子ジャーナル化をするための支援ツールで、J-STAGE独自の形式であるBIB/CITファイルを書き出すものである。このツールはJ-STAGEの標準システムによる学会誌の作成は行わないが、J-STAGEへの搭載は行う学会のために開発されたものである。

Adobe FrameMaker+SGMLの汎用テンプレートも書誌事項作成支援ツールも、どちらもWindows上でしか用いることが出来ない。そこで、WindowsおよびMacOSの両方の環境で利用できるツールに対する要望に応えるために、J-STAGE SGML編集ツールが開発された(科学技術振興事業団, 2002c)。J-STAGE SGML編集ツールは、様々な組版ソフトから書き出された論文のプレインテキスト形式のファイルにSGMLのタグをマークアップするツールである。このツールを使えば、版下作成の工程を変更することなく、また、特定のソフトに依存することなく、SGMLデータを作成できる。またJ-STAGE DTDに特化したソフトのため、SGMLに関する知識をそれほど必要とせずJ-STAGE DTDに則っ

た SGML を記述することができる。

J-STAGE SGML 編集ツールによる電子化プロセス

J-STAGE SGML 編集ツールを用いて学会誌を電子化する場合は、J-STAGE に搭載するために必要な PDF と SGML データを別々に作成する。最初に、組版ソフトから PDF ファイルとテキストファイルを作成し、テキストファイルは J-STAGE SGML 編集ツールを用いて SGML ファイルに変換する。

SGML 化された論文データは J-STAGE の Web アップロード機能を用いて登載される。Web 上で SGML ファイルを put するフォームが用意されており、このフォームから put された SGML ファイルは「抽出プログラム」によって SGML の項目に対応するフィールドとして抽出され、J-STAGE センターの各ジャーナル化のデータベースに登載される。

一方、論文本体の PDF も同様に Web アッ

プロードされ、前述の SGML データとともにデータベースに組み込まれる。

J-STAGE で学会誌を公開する場合は、このデータベースの中から必要なデータが読み出され、公開画面が作成されることになる(図 1)。

J-STAGE SGML 編集ツールのシステム構成

J-STAGE SGML 編集ツールは 4th Dimension(4D Japan, <http://www.4djpn.jp/index.html>) を用いて開発されている。単体アプリケーションとなっているため、利用にあたって他のアプリケーション等は必要ない。論文データの管理は 4th Dimension のデータベースエンジンを用いることで実現されている。論文データの編集は、4D Write と呼ばれる 4th Dimension 用のテキスト編集を提供する外部関数を用いている。他の外部関数として、正規表現パターンマッチングおよび SGML 構文解析機能(パースチェック)を実装している。

パースチェック機能は SP(SGML Parser) エンジン (James Clark, <http://www.jclard.com/sp/index.htm>) を移植することで実現している。Windows においては単体のアプリケーションとしてビルドし、4D から呼び出してパースチェックを行うようになっている。MacOS においては完全に外部関数として実装されており、SP 単独では動作させることは出来ない。

J-STAGE SGML 編集ツールの利用条件

このツールは J-STAGE で開発されたものであり、利用にあたっては利用申込書を科学技術振興事業団に送付する必要がある。利用申込書が科学技術振興事業団に受領されると、ツールの入った CD-ROM が送られてくるので、これを各自のパソコンにインストールして利用することになる。

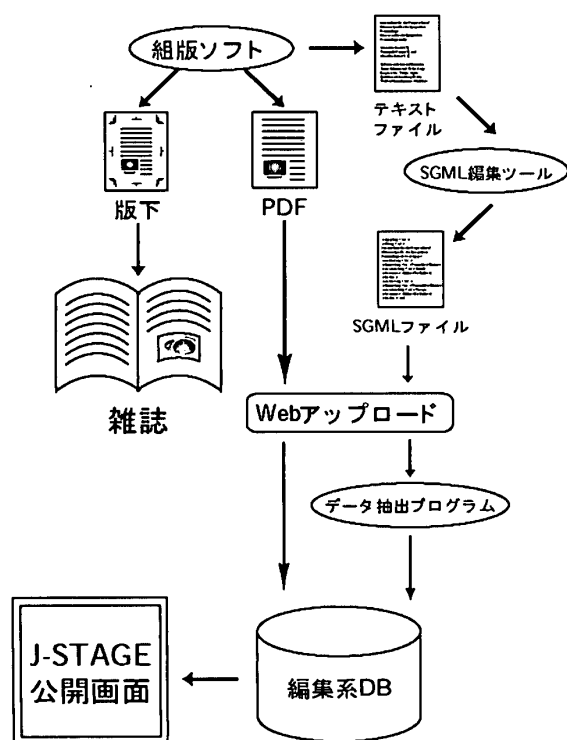


図 1: 雑誌の電子化におけるデータの加工過程

また、初めて Web アップロードの作業を実行する前に、作成した SGML データの J-STAGE センターによる検証を何度か行い、アップロード作業の講習を受ける必要がある。

なお、このツールの動作環境を以下に示す。

Windows

OS: Windows 95/98/NT4.0/Me/2000/XP

CPU: Pentium 90 以上

メモリ: 32MB 以上 (48MB 以上を推奨)

HD: インストール時に約 10MB 必要

Macintosh

OS: MacOS 8.1

CPU: PowerPC

メモリ: 20MB 以上 (32MB 以上を推奨)

HD: インストール時に約 10MB 必要

J-STAGE SGML 編集ツールによる作業フロー

J-STAGE SGML 編集ツールによる SGML 作成の標準的な作業フローは以下のとおりである。

- (1) 学会誌の冊子体の印刷が終了した時点で、印刷会社が組版作業に用いているアプリケーションを用いて、各論文をテキストファイルに書き出す。
- (2) 冊子一冊分のテキストファイルを一括読み込み機能で、J-STAGE SGML 編集ツールに読み込む。
- (3) 編集画面上で、それぞれの論文データに SGML タグを付与する。タグ付けの方法は三種類あるが、作業結果は同じである。
- (4) パーサによる SGML 構文チェックを行う。エラーがある場合は、エラーメッセージを参考にタグ付けの修正を行う。
- (5) 全ての論文全体のデータのタグ付けが完了し、構文チェックにパスしたら、一括書き出し機能により、SGML ファイルを書き出す。

J-STAGE SGML 編集ツールの基本機能

J-STAGE SGML 編集ツールの基本機能としては、論文データの管理、SGML のタグ付け、および、SGML の構文チェックの三つに大きく区分される。以下にそれぞれの機能を個別に紹介する。なお、ここでは、テキストとして書き出された論文と J-STAGE SGML 編集ツールのデータベース上でのレコードとを厳密に区別せず、どちらも論文データと呼ぶことにする。

論文データの管理

J-STAGE SGML 編集ツールは様々な組版ソフトから書き出されたテキストファイルや、SGML ファイルを読み込むことができる (図 2)。フォルダ内のファイルを一括で読み込む機能もある。一括読み込みをする際に、SGML ではエンティティーで表す必要のある & , < , > などは自動的にエンティティー変換することができる。読み込まれたテキストファイルに SGML タグを付与し、フィールド抽出を行うことにより、論文データの情報が内部データベースに反映され、ソートや検索のキーとして用いられる。ソートや検索のキーは巻、号、タイトル、ファイル名などがある。SGML ファイルの書き出しは、一括でも個別でも行える。この時に生成される SGML ファイルのファイル名は書誌情報のファイル名の拡張子を .sgm に変えたものである。また、フィールド抽出された情報はプルーフリストの印刷にも用いられる。

SGML タグ付け

タグを付与する方法は三種類あり、利用者の好みや習熟度に合わせて、それぞれの方法を使い分けることができる。

いずれの方法でも、読み込まれたテキストファイルの中からそれぞれのエレメントに対応

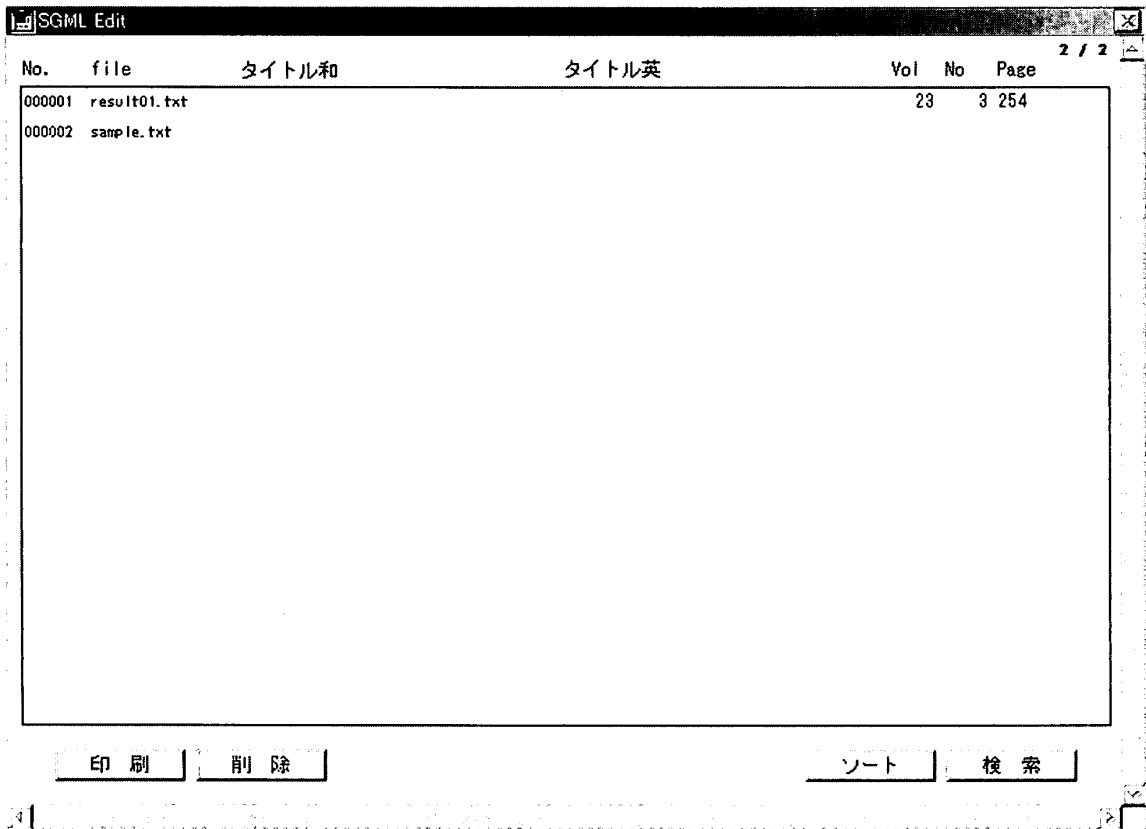


図 2: 論文一覧画面

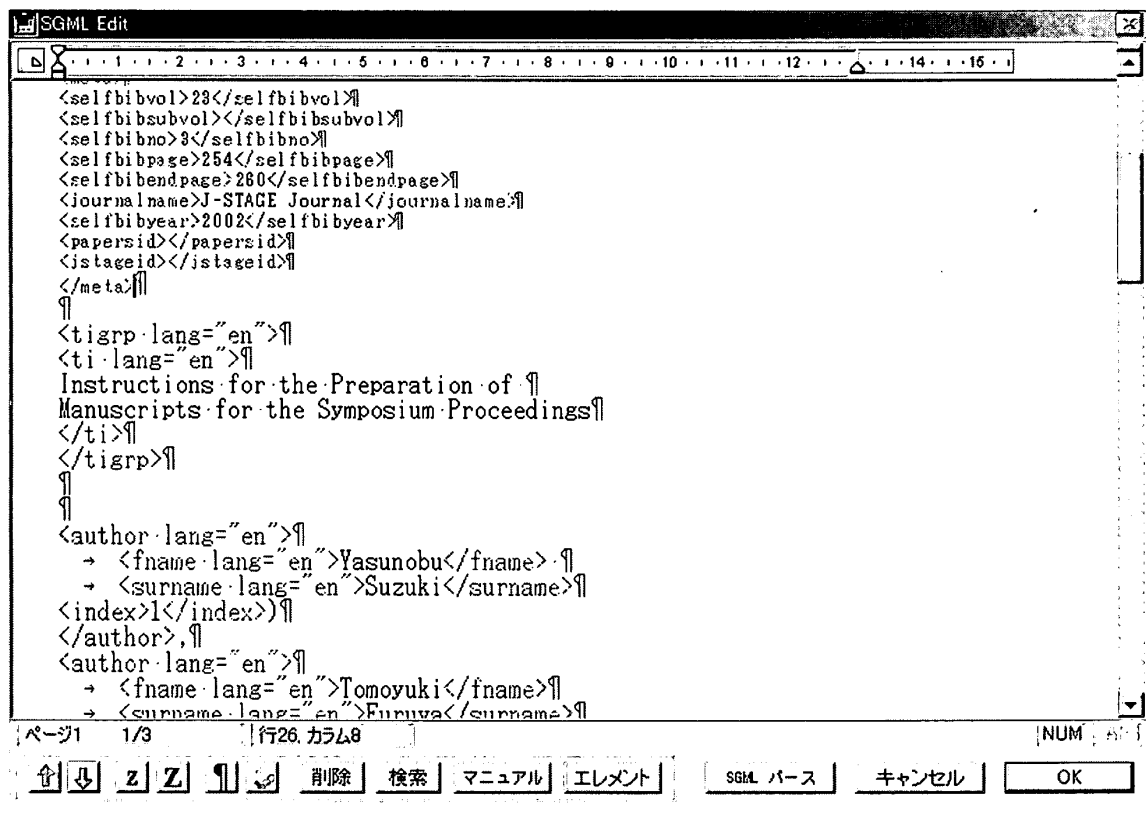


図 3: 編集画面

するデータを指定し、その前後にエレメントに対応するタグを付与することができる(図3)。

第一は、エレメント一覧ウィンドウ(図4)から付与したいタグをダブルクリックする方法である。エレメント一覧ウィンドウには、左側にエレメントの名前、右側には対応する具体的な論文構成要素が書かれている。また、DTDの構造に合わせて、エレメントの項目がインデントされて表示されており、エレメントの親子関係を理解することができる。これらにより、SGMLに関する知識を必要とせずにタグ付けを行うことができる。

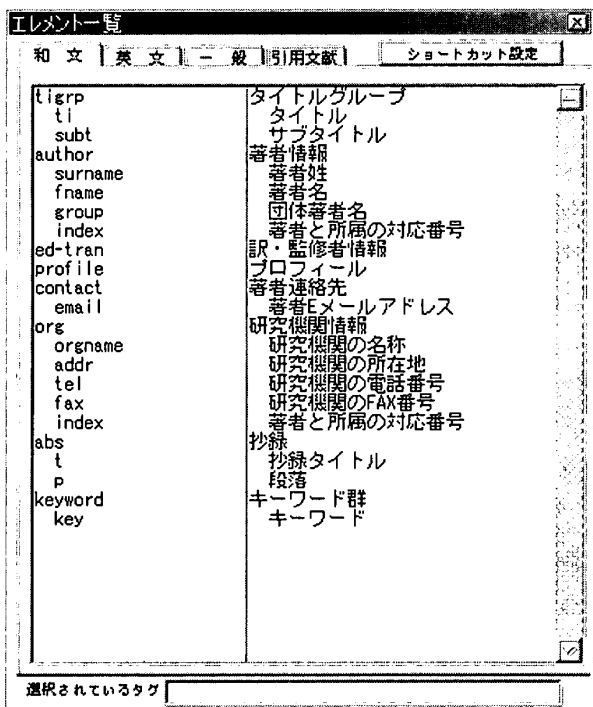


図4: エレメント一覧ウィンドウ

第二は、コンテキストメニューによるタグ付けを利用する方法で、ある程度エレメントの名称が理解できる場合に有用である。Windowsでは右クリック、MacOSではコントロール+クリックにより、コンテキストメニューにエレメントの一覧が表示される(図5)ので適当なタグを選択すると、そのタグが付与されることになる。コンテキストメニューには論文構成要素が書かれていないため、ある程度SGMLに関する知識が必要とされる。しかし、頻繁に使われるエレメントはそれほど多くない

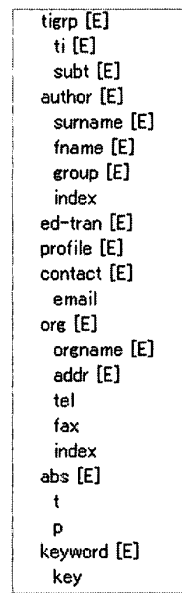


図5: コンテキストメニューによるエレメント一覧

め、エレメント一覧ウィンドウからのタグ付けと併用することで、迅速なタグ付けを行うことができる。

第三は、ショートカットによるタグ付けである。Windowsではコントロールキー+数字の1~9, MacOSではコマンドキー+数字の1~9のショートカットキーに、任意のエレメントを登録することができる(図6)。ショートカットキーを押すと、登録されたエレメントのタグが付与される。このショートカットには、たとえば引用文献のような、繰り返しの多いエレメントを登録することで非常に迅



図6: ショートカット登録ウィンドウ

図 7: 書誌情報入力画面

速なタグ付けが可能になる。

また、直前に入力したタグを再び入力するショートカットも用意されており、繰り返しの多い場合には有効である。

論文のテキストデータ自身には記述されていない可能性のある、巻、号、開始ページ等の書誌情報は書誌情報入力画面(図7)から入力する。この画面で値が入力されると、SGMLヘッダ情報が付加される。

なお、J-STAGE DTDにおいては、本文部分も大部分SGML化することが可能である。しかし、J-STAGEでは論文データの本文部分は全文検索にしか用いられないので、J-STAGEに搭載するためのSGML化においては本文部分の詳細なタグ付けは必要ない。

編集画面においてSGMLのタグは、色付けされて表示される。またSGMLファイルを読み込んだ場合などやタグを手入力した場合には、タグの一括色付け機能により、タグの部分の色付けが行える。

SGML 構文チェック

SGMLのタグ付けが完了したら、パーサにより、J-STAGE DTDに則った正しいSGMLファイルであるか確認することができる。このとき、もし間違いがあれば、パーサがエラー箇所とエラー内容を別ウィンドウに表示する(図8)。エラーの内容は、英語で表示される。エラー箇所をダブルクリックすることで、編集画面における該当箇所にジャンプすることができる。マニュアルにおいて代表的なエラーメッセージについての解説を行っている。

J-STAGE SGML 編集ツールによる作業

前述のように、J-STAGEに搭載することだけを目的とするのであれば、論文の本文については全体を一つのタグで囲んでしまえば良いので、タグ付けすべき項目は限られている。書誌情報については、組版ソフトから書き

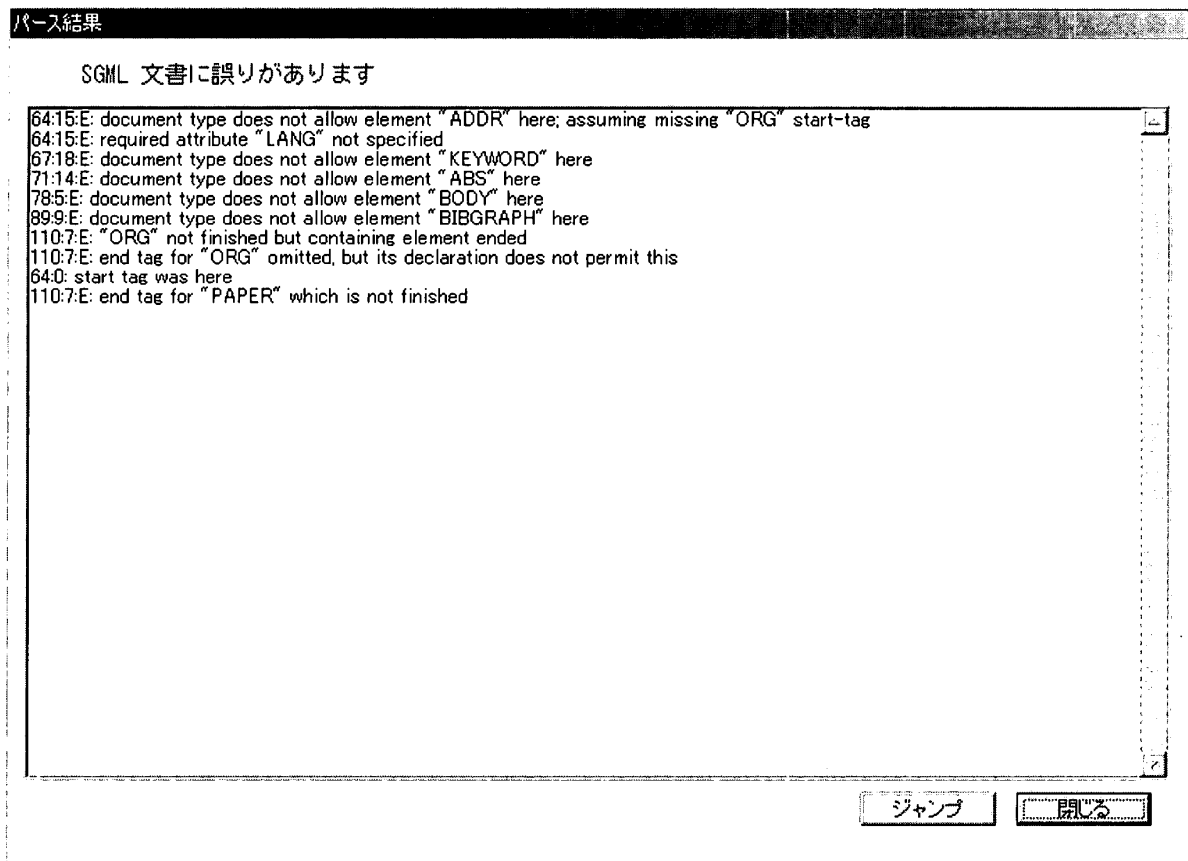


図 8: エラー内容表示ウィンドウ

出したテキストファイルに含まれていないことがある。その場合は、タグ付けに先立って、冊子体の雑誌の目次や奥付のデータを入力しなければならない。

表題、著者に関する情報、キーワード、要旨など論文の記載情報については、もし英文と和文の両方があれば両方ともタグ付けしなければならない。もし、和文か英文の一方しか記載されていない場合は J-STAGE の英語の公開画面において、該当する箇所が表示されないことになる。英語の書誌情報を公開する必要がある場合、冊子を含めた形で項目の見直しが必要になる。

引用文献情報については、平成 14 年 9 月よりサービスを開始する科学技術振興事業団リンクセンターの引用文献自動解析プログラム(説明会での口頭報告)でリンクに必要な情報が自動抽出されるので、引用文献 1 件ごとに区切る為のタグ付けをするだけで、その中の項目毎のタグ付けは省略することも可能であ

る。但し、自動抽出は 100% 完全とは言えないし、タグ付けする段階で不足情報を補うこともあるので、引用文献内部の項目についてもタグ付けすることが望ましい。

筆者らのこれまでの経験では、論文 1 件あたりに要する作業時間は 2~8 時間という大きなバラツキがあった。これは、作業者の SGML の理解度とツールへの習熟度、論文ファイルの完成度と構成要素の複雑さ、著者の数、および、引用文献の数などが影響したものと思われる。また、引用文献 1 件毎の詳細なタグ付けを行わない場合、30 分から 1 時間程度でタグ付けを行える。

考察

J-STAGE に学会誌を掲載したいが、J-STAGE の標準システムを採用することができない学会にとって、学会の従来の編集・印刷システムを変更することなく、J-STAGE

への搭載に必要な PDF および SGML データを、できるだけ少ない労力と経費で作成する手段が求められていた。

J-STAGE SGML 編集ツールは、学会誌の出版後に、J-STAGE に搭載するために必要な SGML データを短時間で作成することを可能とした。ツールに習熟すれば、タグ付けのエラーも少なくなり、作業時間はもっと短縮されるであろう。

ここでは編集委員が SGML データ作成作業を行うことを想定してきたが、今後はアルバイトに作業を委託することも考えられる。あるいは、少ない経費で印刷会社に委託することも可能であろう。

そのためには、J-STAGE SGML 編集ツールにおいて、ウィンドウサイズの変更機能、パースの部分的な実行などの技術面の改良に加えて、エラーメッセージの英文から和文への変更、エラー原因の具体的説明と対策の表示などによって、より使い易いものとするのが望まれる。

このツールの使用に当たって利用者に要請される SGML の知識は基本的な部分のみで十分である。電子ジャーナル化が急速に進む中で、すべての研究者にとっても SGML の基本は常識となっていくことが予想される。そうなれば、将来的にはそれぞれの論文の著者自身が SGML データを作成する方向へ進む可能性も出てきたといえよう。

すなわち、電子ジャーナル化支援ツールの開発・整備によって、学術雑誌の電子化が推進されるとともに、電子化の作業を特定の業者や集団からそれぞれの論文の著者に移行する動きが出てきたといえよう。このことは、電

子ジャーナル化におけるコンテンツの作成を誰が担当するか?という問題に対する答えの一つとなるであろう。

参考文献

- [1] 科学技術振興事業団 (2002a) J-STAGE DTD 解説書. 1.2 版, 72pp.
- [2] 科学技術振興事業団 (2002b) 書誌事項作成支援ツール (MS Access2000/2002 版) 操作マニュアル. 42pp.
- [3] 科学技術振興事業団 (2002c) J-STAGE SGML 編集ツール操作マニュアル. 32pp.
- [4] 白木澤佳子 (2000) J-STAGE の現状と今後の拡張について. 神資研, no.35, pp.17-21.
- [5] 時実象一 (2000) 電子ジャーナルの現状と動向. 情報管理, vol.43, no.5, pp.391-410.
- [6] 吉田幸二・尾身朝子 (2001) J-STAGE: 『科学技術情報発信・流通システム』—電子ジャーナル作成とインターネットによる流通—. TECHNO MARINE, no.861, pp.274-278.
- [7] 吉田幸二・時実象一・尾身朝子 (1999) J-STAGE: 「科学技術情報発信・流通システム」電子ジャーナル作成とインターネットによる流通. 情報管理, Vol.42, no.8, pp.682-693.

(2002 年 7 月 1 日受付)

(2002 年 7 月 3 日採録)

概念構造生成のための階層関係自動抽出法に関する検討

Automatic Extraction of Hierarchical Relationships for Structuralization of Conceptual Structures

森本 貴之[†] 浅川 直輝[‡] 後藤 智範[†] 藤原 譲[§]

[†] 神奈川大学 理学部 [‡] 東京ゲームデザイナー学院

[§] 独立行政法人 工業所有権総合情報館

近年の計算機の高速化、大容量化と低価格化、さらにはインターネットの普及によって情報化が加速度的に進んでいる。今後も増加しつづけるであろう膨大な情報や知識を適切に利用するためには情報の内容に関する、より高度な処理機能が要求される。このような機能を実現するためには、情報や知識の意味を理解しなければならない。そして、そのためには意味関係が表現できる構造化およびその利用のためのアプリケーションあるいはシステムが必要である。このような要求に対して、専門用語を最小単位とした概念構造の生成とその利用のためのシステムの開発を進めている。本研究では、概念構造を生成するためのシステムの一部である階層関係の自動抽出法における問題点の調査について報告する。

The global flow of information is being developed at unprecedented speed. However, users can not sufficiently utilize huge amount of information. Therefore, advanced utilization of contents of information are required. In order to realize such sophisticated utilization, it is necessary to understand meaning and characteristics of information. Therefore, the structuralization is required to represent various semantic relationships among information. In order to satisfy such requirement, we made systems for self organized knowledge resources and information retrieval based on semantic relationships. However, these systems can not make enough structuralized knowledge resources to realize sophisticated utilization. This paper reports problems which are classified and systematized at the method of an automatic extraction of hierarchical relationships which called SS-KWEIC because hierarchical relationships are the basis of semantic relationships in our concepts.

1. はじめに

昨今の計算機の高速化、大容量化と低価格化には目を見張るものがある。また、それに伴うインターネットの普及によって情報化が加速度的に進んでいる。しかしながら、計算機の主要機能は相変わらず数値計算や符合処理に基づくキーワード検索、演繹推論であり、豊富な情報や知識の内容を適切に活用することができない。例えば、Web の Search Engine は大量の情報を取り扱うシステムの例として

挙げられるが、情報の持つ内容が十分に反映されるわけではない。そのため、規模が大きくなるほど検索要求の生成や結果の取捨選択におけるユーザ負担が大きい。そのため、情報の意味に関する高度な機能に対する要求も強く認識されるようになってきている。

このような要求の解の一つとして学習・思考機能が挙げられるが、その実現には情報・知識の持つ意味を理解させる必要がある。そしてそのためには意味関係が表現できる構造

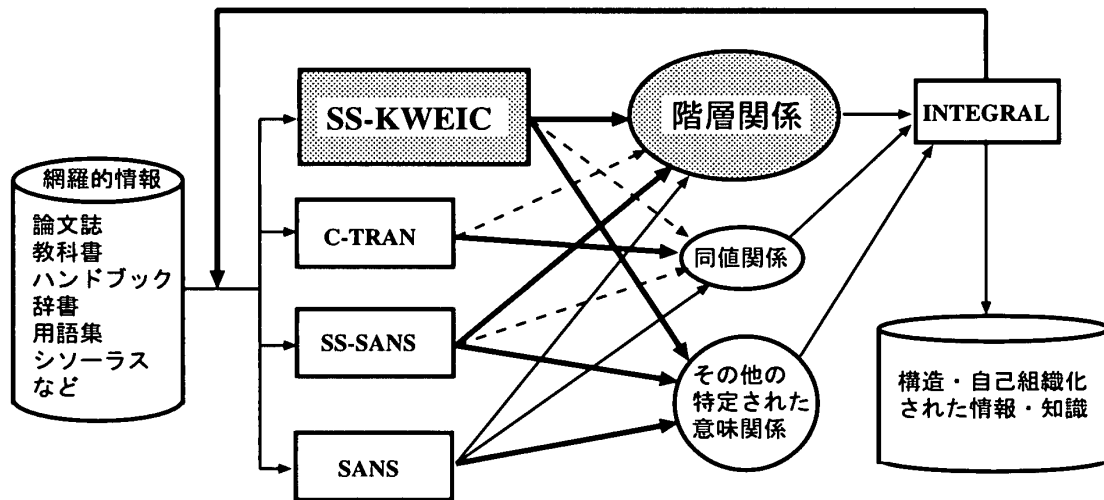


図 1: 情報・知識の自己組織化システム

化が要求される。そこで我々は概念を表現とする最小単位として専門用語を取り上げ、

- 意味関係を表現可能な情報構造モデルの検討 [1][2]
- 意味関係を自動的に抽出、統合、調整するシステム (図 1) の開発 [3][4][5]
- 意味関係に基づいて自己組織化された情報・知識を利用するためのシステムの開発 [6]

を行ってきており、プロトタイプが完成している。

また、学習・思考機能を実装するためには以下の3点が非常に重要である。

- 土台としての大量の情報・知識が必要
- 生成した構造が正確
- 利用目的に応じて構造が適切

これまで意味関係の自動抽出に関しては大量のデータを高速に取り扱うためのシステムの改善を行ってきている [7]。一方、正確かつ適切な構造のための意味関係抽出の改善に関しては一部の例外処理の実装のみに過ぎず [4]、それ以外の問題点に関しては全てを把握できているわけでもない。しかしながら、類推や仮説生成と言った思考機能は意味関係に基づいた構造の解析 (ナビゲーション等) によって実現されるため [1]、間違っただけあるいは不適切

な意味関係および構造は致命的である。そこで、本研究では、階層関係自動抽出法である SS-KWEIC 法における問題点の分類・体系化を行うが、これは学習・思考機能実現のための基礎研究の一つである。

2. SS-KWEIC 法

我々が対象とする専門用語は以下のような特徴を持つ。

- ほとんどが名詞
- 後部分の語基の性質や状態を、前部分の語基が修飾または限定するなどの修飾関係が多い
- 用語は複数の語基を含む複合語であることが多い
- 同じ語基を持つ用語は、何らかの関係を持つことが多い

SS-KWEIC 法は専門用語の構成規則に基づき、複合用語を基本構成用語に分解し、相互の関係を解析することによって階層関係 (包含関係) および関連関係を獲得する方法である (図 1)。SS-KWEIC 法で用いる用語の構成規則を以下に示す。

合成語 ::= 複合語 | 派生語

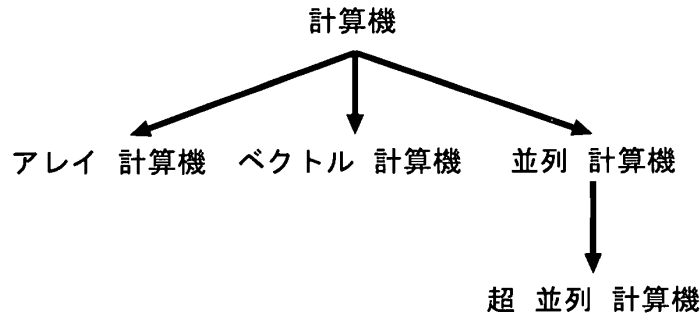


図 2: 階層関係の抽出

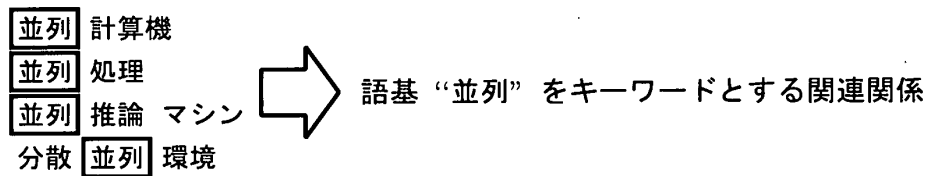


図 3: 関連関係の抽出

複合語 ::= 語基 + 語基 | 語基 + 連結要素 + 語基

派生語 ::= 接辞 + 語基 | 語基 + 接辞

語基 ::= 単純語基 | 複合語基

単純語基 ::= 単純語

複合語基 ::= 語基 + 語基

連結要素 ::= ・ | / | の | な

接辞 ::= 接頭語 | 接尾語 | 数詞 | 量詞

例えば、“並列計算機”は構成規則から“並列”と“計算機”に分解される。そして、上述した専門用語の特徴である、後部分の語基の性質や状態を、前部分の語基が修飾または限定するなどの修飾関係が多いことから、“並列計算機”は“計算機”の下位概念と言う意味関係(階層関係)が抽出できる。同様にしてえられる簡単な階層関係の例を図2に示す。この図では各用語は語基が明確になるように空白で区切っている。本稿では“計算機”と“並列計算機”のように明確な上位/下位概念としてとらえられる場合を特別に包含関係、このような包含関係の集合(例えば図2全体)を階層関係と言いつけることとする。また、同じ語基を持つ専門用語は何らかの関係を持つことが多いという特徴からこのような用語は関連関

係として抽出することもできる(図3)。ただし、本稿では関連関係については言及しない。

以上のように、SS-KWEIC法によって用語の表層情報を利用してシンプルなシソーラスの一種を自動的に作成することができると言える。概念や用語間の関係を記述したシソーラスや辞書としてはすでにEDR^[8]や日本語語彙大系^[9]等があるが、これらは関係自体の精度や利用目的に対する適切さ、さらに最新の情報(用語)への適応と言った面から、知識・情報の構造化の土台として採用するには適切であるとは言えない。一方、SS-KWEIC法においても自然言語を取り扱う以上、階層関係抽出規則が適合しない例も多く、これらが知識・情報の構造化に及ぼす影響は無視できない。

3. 実験

1章で述べたように、本研究の目的は学習・思考機能実現のための前提条件となる、階層関係自動抽出手法における問題点の発見とその体系化である。そこで、SS-KWEIC法による用語間の階層関係自動抽出の検討のための

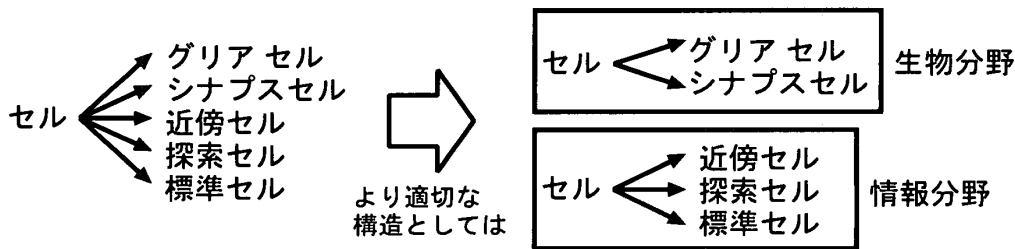


図 5: 用語の多義性

- 英語と日本語の差異
 “**Primitive** Prolog” ⇔ “**原始** Prolog”
- 元素記号
 “細胞内 **Ca**” ⇔ “細胞内 **カルシウム**”

二つ目は“プログラミング言語 Java”と“Java”、“補酵素 チアミン”と“チアミン”のように、上位概念となる語基(群)が前につくことによって(上記の例では“プログラミング言語”と“補酵素”)、後ろの語基(群)の意味を補足する例である。これらは本来は同じ意味の用語であるが、包含関係として抽出される。

4.3. 類義語

類義語による影響はさらに二つに分けられる。一方は用語の中に類義語となる語基を含む場合で、以下にその例を示す。

- “**質疑** 応答 システム”
 ⇔ “**質問** 応答 システム”
- “**表層** 情報” ⇔ “**表面** 情報”

もう一方は用語全体で類義語となる場合で、以下のような例が挙げられる

- “出力 画面” ⇔ “表示 画面”
- “顔 アニメーション”
 ⇔ “表情 アニメーション”

4.4. 異義語

様々な分野の用語を取り扱う場合、同じ用語あるいは語基で意味が全く異なる場合があ

る。典型的な例としては略語が同じ場合が挙げられる。実験では、異なる意味を持つ用語“IP”が情報、医療、工学、化学等の様々な分野で見つかっている。また、人名にも同様の例が存在し、“ド・ゴール”の末尾の語基“ゴール”が見つかっている。

4.5. 多義語

多義的な意味を持つ用語(語基)によって、階層関係中に使用分野や意味の掛け離れた用語が混在することがある。例えば、“セル”は“探索セル”や“グリアセル”“シナプスセル”といったように情報や生物科学といった様々な分野で用いられる用語である。“セル”をrootとする階層関係にまとめられた用語は構造化された知識として利用・応用という観点から見た場合、同じ階層関係でまとめることは適切ではないものと考えられる(図5)。

4.6. General Term

概念が広いあるいは非常に広範囲の分野で使用される general な用語(語基)が用語の末尾(階層関係の root)にくる用語は多義語の場合と同様、適切な階層関係が抽出できるとは言い難い。例えば“ライフサイクル”や“熱サイクル”、“ランキンサイクル”といった用語が“サイクル”を root とする階層関係となる。また、“システム”ではさらに多くの用語がその下位概念として抽出されてしまう。

4.7. 略語

略語による問題には2タイプあり、一つ以上の語基をまとめて略語表現する場合と、一部の語基が完全に省略される場合である。

まず一つ以上の語基をまとめて略語表現する場合の例としては、以下の用語が抽出されている。

- “リポ 蛋白 リパーゼ”
 ⇔ “リポ 蛋白質 リパーゼ”
- “英会話 学習” ⇔ “英語 会話 学習”

このような省略形の語基が用語の最後の語基である場合、SS-KWEIC法では全く異なる階層の集合に構造化される。略語になることで意味が変化する(意図する対象の範囲が広がる)語基もある。“ソフトウェア”の省略形として利用される“ソフト”はこのような例である。

また、用語の中にアルファベットで表現された略語が用いられた場合、それぞれのアルファベットの文字が何を略しているのか判定できず、階層化できないことがある。例として、“intelligent CAI”あるいは“知的 CAI”を意味する“ICAI”は“CAI”との階層関係が抽出できない。

さらに、語基がアルファベット1文字の場合はさらに特殊で、語基は略語、記号、型番等様々なものである可能性があり、それらを識別することは非常に困難である。また、このような語基が用語の最後の語基である場合は階層関係に大きな影響を及ぼす。例えば、図6は本実験で得られた結果の一例であるが、この例では“K”は本来異なるものである。

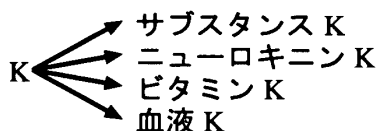


図6: アルファベット1文字による不適切な階層関係

残りの一部の語基が完全に省略される場合

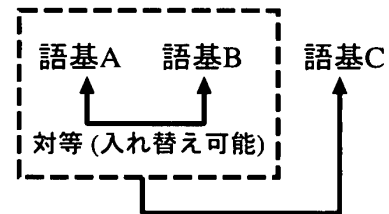
であるが、文字数の長い用語が同じ文章中で何度も使用される際によく現れる。本実験では“応答 内容 生成 システム”と“応答 生成 システム”が抽出されたが、この場合は語基“内容”が省略されている。

4.8. 修飾関係の齟齬

SS-KWEIC法による階層関係の抽出は、前にある語基が後ろの語基(群)を修飾するという規則に基づいている(図2)。そのため、この規則に沿わない修飾・被修飾の関係には対応できない。例えば複数の語基が被修飾語基に対して並列な関係にある場合などがそれに相当する。そのような用語の構成パターン例を図7に示す。



(a) 語基 A, B がそれぞれに語基 C を修飾する場合



(b) 語基 A, B がセット(入れ替え可能)で語基 C を修飾する場合

図7: 修飾関係の齟齬(修飾語基が複数の場合)

本実験では
 “リンゴ酸 アスパラギン酸 シヤトル”
 ⇔ “アスパラギン酸 リンゴ酸 シヤトル”
 が抽出されている。

また、上述とは逆に被修飾の語基が複数ある場合も存在し(図8)、実験で抽出された用語

としては“地質 温度計 圧力計”がある(図8)。しかし、実際には“地質 温度計 圧力計”は“地質 温度計”と“地質 圧力計”をまとめたものである。“圧力計”、“温度計”、“地質 温度計”、“地質 圧力計”、“地質 温度計 圧力計”の5つの用語の階層関係は複雑で、SS-KWEIC法では対応できない。

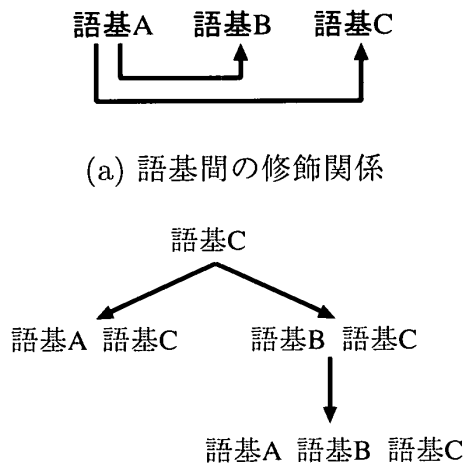


図8: 修飾関係の齟齬(被修飾語基が複数の場合)

4.9. 特殊な修飾語基

前につく語基によって指す意味が大きく変わってしまう場合が存在する。例えば、“ギリシア”と“古代 ギリシア”はSS-KWEIC法では包含関係として取り扱われる。しかし、それらの用語の指し示す全ての「地域・場所」が包含関係にあるとは言いきれない上、“古代 ギリシア”は「場所」ではなく「時代」を意味することもあるため、階層関係として取り扱うには適切とは言えない。

4.10. 接辞

単独では意味をなさない接辞には複数の問題が存在する。“反”、“不”、“非”といった用語の意味を反転させる接頭辞の場合、意味的には並列に位置するものであるが、SS-KWEIC

法では包含関係となってしまふ。また、接尾辞はそれ自体を階層関係の root にした場合、適切ではない階層関係の集合ができてしまふ。さらに、接頭辞と接尾辞の両者を含む場合、その修飾関係は非常に複雑で、接頭辞と接尾辞の組み合わせや間にある語基(群)に依存して変化する。例えば、“高密度化”と“再酸素化”はそれぞれの語基の修飾関係は図9のように異なる。



図9: 接辞を含む語基間の修飾関係

5. おわりに

加速度的に進む情報化において要求される計算機の新しい機能として、情報や知識の持つ意味内容に対する高度な機能の実現に向けて、知識の構造化に関する研究を行っている。本研究は知識の構造化において最も基本となる階層関係の自動抽出における問題点の検討に関するものである。

今後はこれら問題点の対処法の検討および実装を行うことで情報・知識の自己組織化システムの改善を図る。また、この改善したシステムで生成された正確かつ利用目的に適した概念構造を基に、類推や仮説生成と言った思考機能を実装したアプリケーションの開発を進める予定である。

謝辞

本研究では、国立情報学研究所で提供されているテストコレクションNTCIR-1, NTCIR-2の一部を利用した。これらは、科研費報告書および国内学会の提供する学会発表の概要等を利用して作成されたものである。

参考文献

- [1] Y. Fujiwara and Y. Liu., "The homogenized bipartite model for self organization of knowledge and information", IFID, 2(1), pp.13-17, 1998.
- [2] 森本 貴之, 藤原 譲, 情報の構造化とその実装に関する検討, 情報処理学会第 61 回 (平成 12 年後期) 全国大会講演論文集 (3), pp.111-112, 2000.
- [3] T. Morimoto, T. Maeshiro, Y. Fujiwara, "Extraction of semantic relationships among terms to construct organized knowledge resources", Proc. of 1st NT-CIR Workshop on Research in Japanese Text Retrieval and Term Recognition, pp.459-465, 1999.
- [4] 森本 貴之, 藤原 譲, 例外処理を考慮した用語間の階層・関連関係の抽出, 情報知識学会第 8 回 (2000 年度) 研究報告会講演論文集, pp.17-22, 2000.
- [5] 近藤 雄裕, 石川 大介, 池村 匡哉, 杉田 勝彦, 森本 貴之, 藤原 譲, 意味関係抽出による概念の構造化, 情報処理学会第 62 回 (平成 13 年前期) 全国大会講演論文集 (3), pp.199-200, 2001.
- [6] 森本貴之, 近藤雄裕, 杉田勝彦, 石川大介, 池村匡哉, 藤原譲, 構造化された知識を基にした情報検索システム. 情報知識学会第 9 回 (2001 年度) 研究報告会講演論文集, pp.75-80, 2001.
- [7] 杉田勝彦, 近藤雄裕, 石川大介, 池村匡哉, 森本貴之, 藤原譲, 意味関係抽出による知識構造と構築, 情報処理学会第 62 回 (平成 13 年前期) 全国大会講演論文集 (3), pp.197-198, 2001.
- [8] 日本電子化辞書研究所, EDR 電子化辞書仕様説明書.
- [9] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦編, 日本語語彙大系 (全 5 巻), 岩波書店, 1997.
- [10] <http://www-nagao.kuee.kyoto-u.ac.jp/>

nl-resource/juman.html

- [11] <http://research.nii.ac.jp/ntcir/index-ja.html>

(2002 年 7 月 1 日受付)

(2002 年 7 月 5 日採録)

著者紹介

森本 貴之 (正会員)

筑波大学 大学院 博士課程 工学研究科中退.
現在, 神奈川大学 理学部 助手.

知識の自己組織化およびその利用等に興味がある.

情報処理学会会員.

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

浅川 直輝

神奈川大学理学部情報科学科卒.

現在, 東京ゲームデザイナー学院在籍.

後藤 智範 (正会員)

慶応義塾大学 大学院 文学研究科 図書館情報学専攻 博士課程満期退学.

現在, 神奈川大学 理学部 情報科学科 教授.
情報検索の研究に従事.

情報検索理論, 知識構造の視覚化, 自己組織システム等に興味がある.

情報処理学会, 人工知能学会, 日本認知科学会等会員.

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

藤原 譲 (正会員)

東京大学工学部応用物理科卒, 理学博士.

現在, 独立行政法人工業所有権総合情報館理事長.

情報学基礎論とくに情報の特性・構造解析, 意味表現モデル, 自己組織化知識に基づく学習・思考機構およびその応用システムの研究・開発に従事.

ACM, IEEE, ASIS, 電子情報通信学会等会員.

E-mail: fujiwara-yuzuru@ncipi.jpo.go.jp

諸学のなかにおける情報学の位置づけ The Positioning of *Johogaku* in Sciences

平田 周

立正大学大学院経営研究科

高等学校の普通教科「情報」が来年度よりいよいよ実施される。その内容をみると、コンピュータやネットワークのテーマが中心である。高等学校で教える教科のほとんどは、それらが根拠とする諸学が存在する。新教科「情報」はいかなる学が基礎となっているのであろうか。内容からは、強いていえば、コンピュータ・サイエンスであるが、わが国では、コンピュータ・サイエンスが学として認知されていない。「情報」にそれ以上のものが込められているのも事実である。「情報学」がその支えとなるべきだと考える意見は多い。しかし、情報学は、まだ明確な定義を与えられているとはいえない。この問題を考えるにあたり、諸学を二つの異なるタイプに分けて考える、新たなアプローチのスタンスを本稿で提案してみたい。

Among sciences including natural science, social science, and humanities, what role will be played by *johogaku*, information science? The author suggests a new concept that sciences can be classified in types of A series and B series. While A series are many of sciences commonly studied, B series are some special type of sciences such as mathematics, philosophy, law, and linguistics, whose characteristics are combinability with various sciences of A series. *Johogaku*, Information science can be characterized both A and B series. The author named the former “informatics” and the latter “informatology”. This will be useful as an approach in study of the essence of *johogaku*, information science.

1. 高等学校新教科「情報」と情報学

いよいよ平成 15 年度より、高等学校で新教科「情報」が実施され、教科書も決まった。しかし、従来からある教科、たとえば、数学、理科、社会、国語などが、確立されたそれぞれの学を基礎にしているのに対し、「情報」だけは趣を異にしている。本来なら、「情報学」がその裏づけになるべきであろうが、肝心の情報学というのがはっきりしない。「情報」の明確な定義すらできずにいるのであるから、おおよそ無理なことだといえるが。

平成 10 年 7 月に公開された文部省教育課程審議会の答申において、高等学校普通教科として「情報」を取り入れることが提案され、平成 11 年 3 月、学習指導要領案の最終答申が出された。これを背景に、文部省（現文部科

学省）は、独自の調査に加え、教育工学関連学協会連合の小中高一貫情報教育カリキュラム委員会活動、日本学術会議における科学教育に関連する委員会での議論、産業・経済界から提言された 21 世紀に向けての人材教育ビジョン、高等教育機関での高度情報化に向けての教育のあり方、情報処理学会の初等中等教育委員会での議論、科学技術基本計画法などからの提言、さらには文部省中等教育審議会、教育課程審議会などでの議論が行われ、平成 12 年 3 月、これにより高等学校学習指導要領改訂が発表されたのである。^[1]

このなかで、新たに設けられる高等学校普通教科「情報」について、(1) 情報活用の実践力、(2) 情報の科学的な理解、(3) 情報社会に参画する態度の三つを目標にあげている。そ

して、教科として、「情報 A」「情報 B」「情報 C」の3科目が決定された。

「情報 A」コンピュータや情報通信ネットワークなどを活用して、情報を選択、処理、発信できる基礎的な技能の育成に重点を置く。

「情報 B」コンピュータの機能や仕組み、およびコンピュータの活用の方法について、科学的に理解させることに重点を置く。

「情報 C」情報通信ネットワークなどが、社会のなかで果たしている役割や影響を理解し、情報社会に参加する上での望ましい態度を育成することに重点を置く。^[2]

検討の初期段階では、コンピュータ・リテラシーを身につけさせることを新教科「情報」の狙いとする意見が強かった。当時はまだわが国の一般家庭にパソコンはさほど普及しておらず、教師のあいだでも、コンピュータを操作できる人はごくわずかという状況にあった。生徒たちに、コンピュータの操作や利用の知識を与えることが必要と考えられたのだった。しかし、60%近い家庭にパソコンがあり、学校でもパソコンの設置が進んだ環境では、ただコンピュータ関連の理解だけでは、わざわざ新教科「情報」をつくるにはふさわしくないとする空気が関係者のあいだにも広がってきた。「情報 A」「情報 B」では、コンピュータそのものが教える対象となっており、「情報 C」でようやくコンピュータを超えた領域で情報の問題をとらえるような構成になっている。それでも、「情報 C」の内容が、コミュニケーションといっても、やはりコンピュータが中心になっていることに、教科「情報」の本質をみてとれよう。

いまここで、文部科学省が定めた新教科「情報」の内容を分析、批判することは、本稿の意図するところではない。これについては、別の機会に譲る。問題としたいのは、冒頭で述べたように、教科「情報」がいったいいかなる学術領域を背景にしているのかである。

そこで誰もが思い浮かべるのが「情報学」であろう。情報学という学としての概念が存在することは、情報知識学会誌 Vol.10 No.1、

Vol.11 No.1が、小特集「情報学」となっており、この言葉が社会的にある程度認知されているとみてよいであろう。とはいえ、ここでいう情報学に明確な定義が与えられているわけではない。どのあたりまでの学術領域がこれに含まれるかさえも明確にはされていない。まだ試行錯誤の段階にあって、定かにすることが無理、あるいはそれ自体が意味をもたないとする意見もある。

「情報」という言葉そのものが明確に定義されておらず、とりようによってはどこまでも拡大できる内容を包含していることから、情報学の意味を定めることは不可能に近いのかもしれない。しかも、情報という語は、情報経済学とか、情報生物学とか、他の学術分野と結びつきやすい性質もっている。それらの分野をすべて対象とすれば、その範囲は際限なく広がってしまうだろう。

しかし、高等学校の教科で「情報」が正式に教えられることになった以上、それが準拠する学が何であるかを明らかにすることは必要だと思われる。学習指導要領をみるかぎりでは、コンピュータ・サイエンスが基になっているような印象も強い。しかし、コンピュータ・サイエンス自体がわが国では十分認識されておらず、学としても確立したものにはなっていない。「情報 A」「情報 B」の教科書をみるかぎり、コンピュータ技術一般を解説したテキストの延長のようにしか映らないが、新教科「情報」にはそれ以上の思いが込められていることは、学習指導要領の解説からも感じとることができる。

国立情報学研究所の小野欽司は、「情報学とは、情報学基礎理論、情報基盤、ソフトウェア科学、知能科学、情報メディア、社会情報学、情報環境学、バイオインフォマティックス、学術情報研究、教育情報学、経営情報学、応用情報学など、情報に関するあらゆる学問領域をカバーする学問で、英語のinformaticsは情報学を意味する適格な用語と思われる」としている。^[3]

しかし、これでは、情報学の範囲を蓋然的

に示しただけで、ここにあげられている関連分野の種類をみても、まだまだ広く拡大する契機を含んでいる。さらに、「人間と社会との関わりまで追求する、ロジック、コンピューティング、概念形成、システム、あるいはコンテンツを含む応用など、幅広い情報に関する学問領域である」と続けているが、人間以外の生物のコミュニケーションは、情報に含まれないのか、遺伝子情報はどう位置づけるのかなど、疑問は多い。^[4]

教科「情報」は、もっぱら社会的実用性が強調されていて、ここで説明されているような広範かつ高度の概念をふまえているようには思われない。将来、変更されていくのかもしれないが、文部科学省の性格を考えれば、一度定められたものが根本から変更されることは期待しにくいのではないか。

初等・中等教育の正課として「情報」が採用になったいま、あらためて、この問題を考えるよい時機であると考え。

広大な、しかも拡大が続く情報の学術分野から、情報学が何かということ定義することは難しい。とりあえずいまの段階では、情報学の内容や領域を深く追求しようとするよりは、その学としての諸学のなかでの位置づけを考えてみるのがより实际的であり、かつ今後の考察に意義があるのではないかと考える。

2. 諸学を横断的に貫く学の存在

自然科学、社会科学、人文科学といった一般的なくくりでとらえられる諸学は、それぞれに独立的である。以前からいわれるインターディシプリナリーは、学際的とも訳されるように、これら独立関係にあるそれぞれの学の間に出現しうる新たな学術的領域を指し、出現した後は、それもまた一つの独立性をもつようになる。古代ギリシャからの学問の発展は、こうした際限のない専門領域に分化する過程であったと認識することができる。

その一方で、諸学と結びつきやすいいくつ

かの学があることに気づく。たとえば、数学は、数学としての固有の、きわめて専門性の高い研究領域をもつと同時に、それが物理学以外にも、化学や地球物理学、天文学などの自然科学と結びつくのみならず、経済学などの分野でも高度なかたちで取り入れられている。

数学だけではない。哲学も同様な性質をもっている。歴史哲学、科学哲学、法哲学などがその例である。数学と結合した論理学もある。法律学もまた似たような性格もつことから、法学という固有領域だけでなく、経済法など経済学の方々のほか、昨今は生命倫理が関係する生命科学や医学の方面でも重要性を増している。

ここで注目されるのは、これら諸学と広く結びつく契機をもつ学としてここにあげた数学、哲学、法学などには、共通的性格をもっていることが認識される。それは、一種の言語的な性格を共通因子としていることである。数学は科学的記述に適した言語であることは説明の要はないであろう。哲学は論理学のみならず、思考の厳密さを規定するうえでは、欠かせないツールとなりうる。さらに、法律は、規則化するという機能によって諸学と結びつく。

特定の学として規定するにはあまりに普遍化されてしまったが、言語なしには、いかなる学も存在しえないはずである。言語そのものを研究対象とするものとして、言語学があるが、その内容は、諸学と結合する学としては、やや無理もある。適当な名前がないので、ここでは広い意味の言語学としておこう。このなかには、民俗学などの手法も含まれる。言語哲学は、言語学と哲学が結合したものであるが、哲学の一領域としてできあがったと考える。

ここまでくれば、情報学もまた、これら数学、哲学、法学、広い意味の言語学と同列の学として認識されよう。情報が主体的に言語によって構成されるのであるから、当然といえば当然である。しかし、言語が言語学だけではカバーしきれないように、コンピュータ・

サイエンスに、このような任務を負わせることは無理であろう。コンピュータ・サイエンスと情報学の関係も、ほとんど論じられたことはない。

バイオテクノロジーとコンピュータ・サイエンスが結合して誕生したバイオインフォマティクスは、生物学のインターディシプリナリーの一分野と考えるよりは、生物学と情報学が結合して生まれたものだとして解釈したほうがよいであろう。

情報学と諸学の結びつきは、インフォメーション・テクノロジーが今後ますます普及するとともに拡大し、新たな学として新領域をつくっていくであろう。

(注) 生物学と環境学が結びついて研究の新領域を形成するという事は現実にあるが、これは生物環境学といったインターディシプリナリーの領域として考えたい。

3. 情報学の二面構造

前項で述べたように、諸学には、独立的研究対象となるものと、諸学と結合するものとに分けることができる。もちろん、後者の場合も前者の範疇にも入るものであり、独立の研究対象となる。前者のような学を、適当な言葉がないので、とりあえず A 系列、後者を B 系列と呼ぶことにする。[表 1 参照]

A 系列の学同士が結合してインターディシ

リナリーの学を形成するものとは別に、B 系列の学が、A 系列に属する諸学と結びついて新しい分野を生む。誕生した新たな学は、A-A の場合も A-B でも同じものであるかもしれないが、結合のプロセスは違っていると思われる。その特徴は言語性にあると考えられる。それゆえに普遍的な結合性を有している。B 系列の学が、A 系列の学とは異なる性質をもつことが了解される。

平成元年に、文部省が「情報教育に関する手引き」を出したとき、当時すでにコンピュータの普及を予想して、情報教育の重要性が認識され、カリキュラムに取り入れられた。このときは、いわゆる分散型カリキュラムと称し、各教科のなかに情報に関する単元が配置され、数学、理科、社会科などの教科の関連する箇所情報で扱われた。しかし、情報の科学的根拠をもたなかったため、あいまいなものに終わっている。これは明かに、B 系列の考えを暗示させる。

情報知識学会誌や研究報告会で発表される諸研究は、データベース構造を論じるもの、SGML/XML などコンピュータのマークアップ言語など、情報学の専門領域を扱うものが中心をなす一方、歴史学やバイオテクノロジー分野での応用にかかわるものも多く目につく。これらは、いずれも A 系列に属するものである。今後、これらの分野における諸研究はますます多彩になり、分化を遂げるとともに、諸

表 1: 諸学の二面構造

A 系列 \ B 系列	自然科学					社会科学						人文科学						
	物理	化学	生物	地質	数学	情報	経済	政治	社会	法律	心理	人類	歴史	哲学	文学	芸術	言語	宗教
数学	◎				◎	○	○							○				
哲学			○		○			○	○	○			○	◎	○	○	○	○
法学			○			○	○	◎	○	◎				○				○
言語学						○			○		○	○	○	○	○		◎	
情報学			○		○	◎	○		○				○		○		○	

(注) ○ はある程度の関連、◎ は強い関連

学との関連を強めていくであろう。このような幅広い諸学と結びつく契機をもっているのは、情報学がB系列の性格のものであることに理由を求めることができよう。

それゆえに、数学、哲学、法学のように、それそのものの本質を探求する研究、すなわちB系列としての情報学というアプローチが求められる。この場合、A系列にある情報関連の諸学から帰納的に情報学の本質を探ろうとすることには困難が多いと予想される。もし情報学の本質を極めようとするならば、B系列からみるスタンスが必要である。同様にB系列である数学、哲学、法学、あるいは広い意味での言語学は、長い年月をかけて行われてきたA系列から帰納された結果でもあるが、やはりその本質から探るB系列からのアプローチが根底をなしてきたと考えられる。情報学が、まだ歴史的にみて日が浅いことは事実であるが、B系列の学として、同じような立場に立つべきである。

文部科学省が、以前、分散型カリキュラムとして情報に関連教科で教えようとしたことの失敗は、明らかに、このB系列としての情報学研究の基盤のなさに原因を求められるはずである。そして、このたび新教科として定められた「情報」の内容が、ことごとくA系列に属するものになっている点は、B系列としての情報学研究の不在を証明するものになっていると思われる。

4. 情報知識学会に期待されるもの

情報学が何かを究めようとするならば、B系列の視点からの研究が重要である。これは、A系列の研究価値を低めるものではもちろんない。研究の数としては、圧倒的にA系列のものが優勢になってしかるべきである。しかし、ここで扱われる研究課題は、情報処理学会など、他の関連学会がテーマとするものと、オーバーラップしてしまうことは避け難い。それはそれで誤りではない。大いにその範囲を広げるべきである。しかし、情報知識学会

が、情報知識学会としてのユニークさ、あるいはアイデンティティを確立しようとするならば、B系列からの研究を刺激することにこそ意義があるのではないだろうか。

前掲の国立情報学研究所の考えも、すべてA系列からのアプローチにとどまっているかのようにみえる。

海外でも、この問題が重視されているわけではない。コンピュータ・サイエンス、あるいはインフォメーション・テクノロジーという言葉は、学校教育でしばしば用いられる。ほかには、インフォメーション・スタディとか、コンピュータ・エデュケーションなどといった言い方もある。しかし、情報学という概念はあまりないようである。情報学にあてる英語としてinformaticsをあてるのが普通である。bioinformaticsという言葉が定着していることから、これは正しいようにも思われる。しかし、本稿では、あえてこれまで英語として使われていないinformatologyという言葉をあてたい。-ticsという接尾語よりも-ologyのほうが範囲は広い。A系列の「情報」をinformatics、B系列の「情報」をinformatologyと名づけてみた。

B系列からの研究が、A系列を軽視するものであったり、純粹に演繹的な立場を固執すべきでないことも確かである。困難は予想されるが、B系列における研究から情報学の本質をとらえる努力が必要である。しかし、いずれのサイドに立とうとも、情報学が、その問題の設定や方法論において、あまりに未熟であることは否定できない。ここにこそ、情報知識学会が取り組むべき課題と任務があるのではないかと考える。

参考文献

- [1] 岡本敏雄他,「教職必修 - 情報科教育のための指導法と展開例」, 実教出版, 2002
- [2] 文部省,「高等学校学習指導要領解説」情報編, 開隆堂出版, 2000
- [3] 小野欽司他,「情報学とは何か」, 国立情

報学研究所監修, 丸善, 2002

- [4] 平田周, 「可能態としての反情報概念」, 情報知識学会誌, Vol.11, No.1 (Apr. 2001)

著者紹介

平田 周 (正会員)

国際基督教大学教養学科卒

三菱レイヨン(株)、Booz Allen & Hamilton、科学産業開発(株)、慶應義塾大学 SFC 研究所員等を経て、現在、立正大学大学院経営研究科講師、IT 経営研究所代表

E-mail: LDG01371@nifty.ne.jp

~~~~~

## 情報関連新刊図書一覧 2002 年上期 (1 月~6 月)

- 看護情報学への招待 キャサリン・J・ハンナ他編集 中山書店 2002・1 2800 円
- IT 革命の虚構 緑風出版編集部著 緑風出版 2002・1 2000 円 [1800 円]
- ユビキタス・ネットワークと市場創造 野村総合研究所著 野村総合研究所広報部 2002・1
- 教科書情報と社会 小暮仁著 日科技連出版社 2002・1 2000 円
- 情報消費社会の経営戦略 佐藤典司著 経済産業調査会 2002・1 2000 円
- ランダム・ウォーク - 乱れに潜む不思議な現象 津野義道著 牧野書店/青雲社 2002・1 1500 円
- インターネット時代の著作権 岡本薫 全日本社会教育連合会 2002・1 1600 円
- 日米知的財産権問題アップデート 2002 アイ・エル・エス出版 2002・1 18000 円
- 情報化と社会心理 林茂樹編著 中央大学出版部 2002・2 3500 円 [429 円]
- 全産業を変革する IT フィールド大研究 就職ジャーナル/Tech.B-ing 共同編集 Recruit 2002・2
- ソフトウェアのための基礎数学 鑰山徹著 工学図書 2002・2 2300 円
- 言語としての民族 - カウツキーと民族問題 相田慎一著 御茶の水書房 2002・2 9500 円
- 情報デザイン 情報デザインアソシエイツ著 グラフィック社 2002・2 2800 円
- 情報メディア入門 高橋参吉編著 実教出版 2002・2 1500 円
- 情報学とは何か 小野欽司他著 丸善(情報学シリーズ3) 2002・2 2400 円
- 情報文明学の構想 吉沢英成編 以文社 2002・2 3500 円 [2667 円]
- くらしの中に IT 文化を くらしのりサーチセンター編集 くらしのりサーチセンター 2002・2
- 情報エネルギー化社会 ポール・ヴィリリオ著 新評論 2002・3 2400 円
- コミュニケーションを科学する 石井健一郎編著 NTT 出版 2002・3 2200 円
- 情報政治学 鶴木真著 三嶺書房 2002・3 2900 円
- メディアとコミュニケーションの教育 水越敏行・情報コミュニケーション教育研究会編著 日本文教出版 2002・3 1500 円
- Linux の哲学 岩谷宏著 ソフトバンクパブリッシング 2002・3 2000 円
- 環境としての空間 遠藤薫編著 アグネ承風社 2002・3 3300 円
- IT 革命と新世紀の社会 南窓社 (横浜商科大学公開講座) 2002・3 2800 円
- 情報と意思決定 田中伯知著 自由社 2002・3 2300 円
- 高度情報化社会の諸相(改訂増補版) 折笠和文著 同文館出版 2002・4 2900 円
- 量子情報科学の基礎 広田修著 森北出版 2002・4 3600 円 [1600 円]
- 基礎情報科学 東洋英和女学院大学 2002 年版 有田富美子他著 日本教育訓練センター 2002・4

- ケースブック情報公開法 宇賀克也著 有斐閣 2002・4 「2000円」  
電子政府時代の政府調達改革 経済産業省情報処理振興課編 コンピュータ・エージ社 2002・4  
ビジネスと情報 池内健治編著 実業出版 2002・4 2000円  
IT社会の法と倫理 サラ・バーズ著 ピアソン・エデュケーション 2002・4 2800円  
コミュニティ・ソリューション 金子郁容著 岩波書店 2002・4 2300円  
ITとマネジメントの接点 岡本英嗣著 白桃書房 2002・4 2800円  
CNCネットワーク革命 小笠原泰他著 東洋経済新報社 2002・5 1800円  
情報科教育のための指導法と展開例 岡本敏雄他編著 実教出版 2002・5 2000円  
IT&バイオ入門 杉野昇編著 オーム社 2002・5 2300円  
学術情報と知的所有権 名和小太郎著 東京大学出版会 2002・5 3800円  
知識と推論 新田克己著 サイエンス社 2002・6 1500円  
IT社会総合データブック 2002 生活情報センター編集 生活情報センター 2002・6 14800円  
情報システム監査 吉田洋著 税務経理協会 2002・6 3000円 「1800円」  
デジタル・エコノミー 2002/03 米国商務省著 室田泰弘編訳 東洋経済新報社 2002・6

(注) 図書館流通センター刊『週刊新刊全点案内』(1256-1279号)を参考にした。

コンピュータ関連の技術・操作関連のものは除外した。

初心者入門的なもの、ビジネス関連の強いものは除外したものもある。

作成者 平田 周 2002/6/28

## 出版コンテンツ配信ビジネスの現状と課題 Publishing Contents

深見 拓史

(株) 廣濟堂 (情報知識学会・評議員)

### はじめに

本年の情報知識学会総会で藤原鎮男会長がその挨拶の中で、コンテンツとエンドユーザとの新たなコンテンツ流通の仕組み作りの創生が重要な課題であると指摘された。

商業出版におけるコンテンツ配信ビジネスの現状と課題について述べてみたい。特にブロードバンド時代を迎えて、コンテンツ配信ビジネスの状況がどのように変化し、商業出版との関わりはどのようになっているのかについて触れてみたい。

### コンテンツ配信ビジネスの現状

インターネット人口が2002年6月時点で7000万人に達したと報道されている。これらのインターネット人口の内、大手プロバイダーであるニフティーやビッグロブ、ソネットなどを仲介してコンテンツを配信する、いわゆるコンテンツシンジケートといわれる凸版印刷のBitwayもある。コンテンツホルダーとユーザの間を仲介し、不足するコンテンツの仲介者となり、利便性をあげようとするものである。これらは既存の雑誌などからコンテンツをネット用に変換もしくは再編集している。着実に伸長していると聞くものの爆発的とは言えないようである。また最近マンガのコンテンツを紙面イメージでインターネット専用ブラウザでサービスする試みがイーブックイニシアティブ社からも提案されており新しい試みとして注目を集めている。また出版社である講談社の「Web 現代」など出版社独自のサイトを持ち配信しているものもある。しかし商業出版社が独自に立ちあげるフルコ

ンテンツ・サイトは未だ十分ではなく従来の紙を中心としたメディアから独立したメディアとして認知されるに到っていないというのが現状であろう。

### ビジネスモデルの崩壊

コンテンツといっても様々であるが、筆者にとって商業出版におけるコンテンツ配信とは有料コンテンツ配信を意味する。無料で配信されるもの、たとえば企業がホームページ等で紹介する会社案内や商品カタログ紹介、さらには政府や自治体が情報公開するものもある。商業出版コンテンツとは、本来有料で配布されるべきもので、著作者や編集者さらには出版社などがビジネスそのものとして企画・編集・制作・配布することによって成立するモデルのなかで存在しうるものである。当初、広告収入モデルで成立させようとしたがそろそろ限界にきているのではないのかと思う。インターネットメールにしろ、ネット配信コンテンツにしろ有料が前提で成立できるような、著作権保護、課金、高速安価な通信、モラル維持されたコンテンツなどのインフラストラクチャーが求められている。そうでなければ良質のコンテンツは、いつになっても配信されそうもない。現在あるビジネスモデルは、見なおすべき時期に来ているのではないのかと思う。

### 有料配信の芽生え

ニューヨークタイムズやウォールストリートジャーナルなどの著名な新聞社が有料コン



テック配信に踏み切ったとのニュースが届いたのは最近のことである。経済原則を無視してきた今までの新聞コンテンツ配信モデルが行き詰まってきたともいえる。考えてみればごくごく当然のことである。新聞制作にかかる費用分担をインターネット読者には要求しないというのは不自然なことである。紙をベースにしたニュースよりもネット配信の方が高い鮮度であることは自明である。また良質なニュースを保証するためには、それなりのコストをかけなくてはならないからである。従来の無料コンテンツ配信ビジネスモデルがもはや崩壊していることをコンテンツホルダー、編集加工者、ビューアのソフト・ハード開発者、ネットワークインフラ保有維持者などすべての関係者が協議すべき時がきている。携帯電話の着メロ配信などですでに実現しているビジネスモデルなどが商業出版物においても成立するようにしなくては、インターネット配信での良質でバラエティーに富むコンテンツはいつまでたっても出現しないことになるのではと恐れている。

## ビューアの開発状況

現在のところ、人々はディスプレイで文章や図形、写真、絵画といったコンテンツを見たり読んだりすることに、十分な満足を得ていない。特に500年来慣れ親しんできた書籍や雑誌など紙面上で情報を手に入れてきた人々にとってディスプレイ上で見たり読んだりすることが必ずしも人にやさしい、とはいえないからである。そのような視点から日本語独特の縦書きや書体へのこだわり機能なども開発されている。印刷との連続性・親和性を保証しようとする試みがアドビ社のPDFであったりする。アメリカでは紙の書籍がなくなるであろうとの考えを持つ人達も現れて、ハードウェアとしての新たなペーパーライクディスプレイの開発やe-Paperなどの話題はいつもにぎやかである。電子書籍ビューアとして携帯型の読書専用端末を開発する動きや、より紙

面上のイメージに近づけるためのグレイフォントを開発するソフトウェアベンダーも存在する。e-Bookリーダのソフト開発も盛んである。限りなく紙のフレキシブル性を保ちかつ軽量で簡便なものを狙い商品開発は熾烈である。しかし、果たして紙を超えられるであろうか？イスラエルのベンチャー企業、APRION社ではインターネットで情報を検索し、その内容を冊子体の書籍にしてしまうという新たな試み「BookNet」も始まっている。世界最大の印刷機材展 drupa 2000 で参考出品されたが、インターネットで情報検索したあと、本文はもちろん表紙まで印刷し、さらには製本されて出てくるというコンセプトである。

## 課題の整理

現在までのところインターネットでは、優れたもののコンテンツは流通していないというのが、私の認識である。たしかに何百万冊とある書籍のなかからお目当てを探しだすのにインターネットは便利であるし、次々とアップデートされていく情報に関して、あるいはホットな話題にインターネットは優れている。また広告媒体としてのインターネットも次第にその真価を発揮しつつある。しかし良質なコンテンツを果たして入手できるであろうか？現在ではノーといわざるを得ない。なぜか？会員登録やパスワード入力、ダウンロードの手続きの煩雑さ、著作権問題や、その防止策としての電子透かし技術など解決すべき課題が山積みされているからである。また一方政府・自治体などの情報公開や電子図書館による有料・無料コンテンツの区分け基準の明確化などもこれからの大きな課題である。しかし最大の課題は良質なコンテンツが有料で配信され受信できる環境やインフラストラクチャーの整備こそ、もっとも急がれる課題であろう。情報発信する側にも受信する側にも大いに問題ありといえる。コンテンツ配布を業とする商業出版社にとってビジネスサイクルがバリアなく廻るようなインフラこそが今求められ

ている。

## 電子認証プラットフォーム提供の試み

一般的なパソコンでのインターネット利用に対して、(株)廣濟堂は日本経済新聞5月13日付けでコンテンツをネットで有料配信するシステムを開発した。キーポイントは電子認証プラットフォームの提供であり、有料会員制の仕組み提供の試みである。ブロードバンド時代を迎えて、ストリーミング動画配信サービスとの組み合わせで発表した。これらの試みが、また一つ次のステップへの飛躍となればと考えている。

## メディアの変遷

### —技術がコンテンツを変え、やがてメディアとなる—

インターネット・メディアが、どの時点でペーパーメディアを追い越すのかあるいは永遠にペーパーメディアを超えることが出来ないのか、もしくはコンテンツによる棲み分けが進むのか、についても現時点では不明である。人間の思考が文字や画像を通じてどのようなインターラクションで行われているのかについての洞察も必要であろう。この新しいメディアはラジオやテレビといった時系列に流れてくるメディアとも異なる。いつの技術もそうであったように、インターネットを中心とする技術がブロードバンド時代到来で激しく変化進歩しようとしている。口伝による情報の伝達から写経になり、やがてグーテンベルグ活字発見と印刷技術によって、メディアが広がりを見せ、連続映像を実現する技術が演劇から映画を生み出し、電波による伝達でテレビが出現したように、インターネットもそのメディア特性が活かされていく。

P. F. ドラッカーによれば「IT(情報技術) 革

命のインパクトは現れはじめたばかりである。しかし何がe-コマースに乗り、何がe-コマースに乗らないはわからない。流通チャネルとはそういうものである。」

書籍や雑誌に変わる新しいコンテンツが生み出される苦しみかもしれない。それにしても、読者が代金を払い、探してみたい、見てみたい、読んでみたいと思わせるような良質なコンテンツとなることが必要ではないのだろうか? 新しい技術がコンテンツを変化させ、インターネットに最適で、優れもののコンテンツが発信・受信されるようになって始めて新たなメディアとなる。このことを期待したい。

## 参考文献

- [1] 深見 拓史, 「商業出版におよぼすインターネットの影響」, 第8回(2000年度)情報知識学会・研究報告会講演論文集 2000年5月
- [2] 深見 拓史, 「商業出版におけるコンテンツ配信の課題」, 第9回(2001年度)情報知識学会・研究報告会講演論文集 2001年5月
- [3] 深見 拓史, 「出版コンテンツ配信ビジネスの現状と課題」, 出版学会, 2002春季研究発表会論文, 2002年4月
- [4] P. F. ドラッカー著, 上田 惇生訳, 「ネクスト・ソサエティ」, ダイアモンド社, 2002年5月
- [5] 凸版印刷(株) Bitway  
URL <http://www.bitway.ne.jp/>
- [6] (株)イーブックイニシアティブジャパン  
URL <http://www.ebookjapan.ne.jp/>
- [7] 講談社 Web 現代  
URL <http://kodansha.cplaza.ne.jp/>
- [8] Aprion Digital Ltd.  
URL <http://www.aprion.com/>

## 専門用語研究部会参加へのお誘い

専門用語研究部会部会長 細野 公男

専門用語研究部会（1988年専門用語研究会の名称で発足）の目的は、「専門用語に関する研究の促進、情報交換の場の提供、会員相互の交流を通じて、専門用語に関する研究や技術の向上に貢献すること」です。わが国においては、明治維新以来科学技術の先進諸国に伍するため、専門領域ごとに用語集の作成に多大の努力を払い、それぞれの分野での研究活動および情報交換が行われてきました。さらに1980年代以降は、情報化社会の進展やITの国家事業化に伴って、学術活動の範囲が学際的かつ国際的になり、また電子媒体による情報交換ビジネスが地球的規模で実施されるようになりました。そして情報交換の基盤である専門用語のもつ役割が改めて認識され、専門用語の研究および実務は新たな展開をみせております。

こうした状況を踏まえて専門用語研究部会の研究分野は、以下のように広範囲にわたっています。

- ① 専門用語の概念、概念と用語の関わり合いに関する研究
- ② 専門用語集編纂技法の研究
- ③ コーパスなど、用語辞書の編纂に関する研究
- ④ シソーラスなど、データベースの構築・検索と係わる手段に関する研究

そして、こうした領域での研究成果は、専門用語シンポジウムの開催、学会誌への投稿、国際会議（ISO/TC37、EAFTerm（東アジア専門用語フォーラム）、TKE など）への参加などを通じて、公表してきました。

Web上にみられる多言語情報の著しい増大によって、言語を資源の一つとみるようになった世界的な流れが専門用語の研究の必要性をさらに高めることは、確かであります。当学会会員各位が専門用語研究の目的・意義をご理解いただき、本部会へ参加されるようご案内申し上げます。また、研究課題の提案、共同研究の推進など、部会活動促進のための御意見があれば、部会事務局までお寄せ下さるようお願いします。

なお、第15回専門用語シンポジウムを下記の要領で開催しますので、こちらの方にも奮ってご参加ください。

開催日時： 2002年12月14日（土）

13:00 - 17:00 シンポジウムおよび一般講演

17:00 - 17:30 専門用語研究部会総会

18:00 - 20:00 懇親会

開催場所： 慶應義塾大学三田キャンパス（西校舎513番教室）

シンポジウム主題：用語政策／用語施策の現状と将来構想

一般講演： ISO/TC37 国際会議報告 ほか

一般講演を下記の要領で募集しております。

提出期限： 2002年8月30日

提出先： 111-0051 東京都台東区蔵前3-1-10 株式会社システムソフト内

情報知識学会専門用語研究部会事務局

E-mail: osada-t@systemsoft.co.jp

提出資料： 標題（専門用語に関するもの）

報告者（氏名および所属）

## CODATA 部会関連の活動状況報告

平成 14 年 5 月 1 日～平成 14 年 7 月 15 日

CODATA 部会部会長 岩田 修一

CODATA 部会研究会、Task Group の主催するワークショップ他を通して多くの議論に参加し、CODATA 部会としての役割を検討してまいりました。以下、今後の重要課題を提案致します。部会研究会、国内・国際会議等の機会に議論を深め、具体的な活動への展開を計りたいと考えております。積極的なご参加をお願い致します。

- 〈組織-関係性〉：各種データ連携（国際、国内、異分野、産官学、世代間、市民他）のための実施体制、情報インフラの構築と拡充の方策。
- 〈価値-ミッション、時代性〉：データ活用による未来志向の価値創出の範例、データの誤用（あるいは無知）による失敗事例の集積と活用。科学技術倫理、知財権、競争と共創。

### 【CODATA Task Group 活動状況】

- Data on Natural Gas Hydrate：5 月 20 日、24 日に横浜シンポジウムで開催の International Conference on Gas Hydrate の機会に関係者が集まり、今後の研究協力体制について打合せを行った。
- Fundamental Constants：6 月 22 日オタワにある NRC(National Research Council) Canada において開催。
- Data Sources in Asian-Oceanic Countries：7 月 5 日、6 日に韓国 Daejeon の KISTI(Korea Institute of S&T Information) においてワークショップを開催。
- Data, Information and Visualization：7 月 11 日、12 日にパリの国立自然歴史博物館においてワークショップを開催。

### 【CODATA 部会研究会】

第 1 回 平成 14 年 5 月 15 日「物質・材料現象のヴァーチャル化構想と活動」

- (1) 仮想実験プロジェクトの構想（二瓶正俊）
- (2) 仮想実験技術を活用した材料設計統合システムー仮想実験プラットフォームの開発（西川宜孝）
- (3) 材料情報分野の動向報告と戦略（岩田修一）

第 2 回 平成 14 年 7 月 15 日

- (1) CODATA TG Fundamental Constants の活動（藤井賢一）
- (2) 鳥類データベースの構築と活用（神山和夫）
- (3) エネルギー環境データの活動（芝池成人）

情報知識学会主催・第7回 SGML/XML 研修フォーラム  
**「XML と Web サービス入門」**

**[開催趣旨]**

いま、次世代のシステム・アーキテクチャーとして「Web サービス」が注目を集めている。これは、XML と SOAP(XML のメッセージング機能) によって疎結合的に実現されるアプリケーション群とされ、社会的なシステムの統合技術として、企業間取引などを抜本的に効率化するものとの期待がある。

こうした情勢を踏まえて、本フォーラムでは、改めて XML の基本を確認し、この基礎のうえに構築される Web サービスについて、その考え方と関連技術の動向、適用分野を含めて、その全体を見渡すことを目標とする。

**[概要]**

開催日：平成 14 年 10 月 22 日 (火)～23 日 (水) の 2 日間

会場：グランドヒル市ヶ谷 (東京/JR および地下鉄市ヶ谷駅より徒歩 3 分)

定員：100 名

参加費：一般 28,000 円 (入会申込者は今年度年会費込み 20,000 円)、  
 リピーター 18,000 円、地方自治体 18,000 円、協賛団体会員 18,000 円、  
 情報知識学会正会員・法人会員とも 14,000 円、情報知識学会学生会員は無料、  
 一般学生 (非会員)3,000 円 (ただし、当日入会申込者は入会金 1,000 円および  
 今年度年会費 4,000 円が無料となる)。

協賛団体：約 20 団体、現在交渉中

プログラム：最新版はホームページ (<http://www.jsik.jp/>) に掲載。

申し込み：いつでも可。郵便番号、住所、氏名、電話・FAX 番号、E-mail、および  
 参加費金額を下記の事務局へ E-mail、FAX または郵送で連絡。

〒110-8560 東京都台東区台東 1-5 凸版印刷 (株) 内 情報知識学会事務局  
 E-mail: LDE01013@nifty.com FAX: 03-3837-0368

**[プログラム]**

|             |                      |                                           |
|-------------|----------------------|-------------------------------------------|
| << 第 1 日 >> | 2002 年 10 月 22 日 (火) |                                           |
| 9:55～10:00  | 開会挨拶                 | 根岸正光／国立情報学研究所教授<br>／SGML/XML 研修フォーラム実行委員長 |
| 10:00～11:30 | XML 入門               | 原田隆史／慶大助教授                                |
| 11:30～12:30 | { 昼食 }               |                                           |
| 12:30～14:00 | Web サービス             | 浦本直彦／日本 IBM 東京基礎研究所<br>／国立情報学研究所客員助教授     |
| 14:00～15:00 | セマンティック・ウェブ          | 飯島正／慶大教授                                  |
| 15:00～15:20 | { 休憩 }               |                                           |
| 15:20～16:20 | e-キャンパス              | 海田茂／ネクストソリューション (株)                       |

<< 第2日 >> 2002年10月23日(水)

|             |                 |                                          |
|-------------|-----------------|------------------------------------------|
| 10:00~11:00 | XMLの標準化.....    | 小町祐史/松下電送システム<br>/ISO/IEC JTC1 SC34 委員長  |
| 11:00~12:00 | トピック・マップ.....   | 内藤求/シナジー・インキュベート取締役                      |
| 12:00~13:00 | { 昼食 }          |                                          |
| 13:00~14:00 | XML署名.....      | 杉山高弘/NECソリューションズ<br>/ニューメディア開発協会XML署名委員長 |
| 14:00~15:00 | 行政電子化.....      | 地方自治情報センター(講師交渉中)                        |
| 15:00~15:20 | { 休憩 }          |                                          |
| 15:20~16:20 | XMLによるBtoB..... | (講師交渉中)                                  |

以上

### <情報知識学会会員の皆様へ>

今年で第7回を迎えますが、一般のかたはもとより、情報知識学会会員も多数参加されています。お知り合いのかたもお誘い合わせのうえ、大勢ご参加ください。情報知識学会会員の紹介で参加申込をされるかたは、参加費28,000円を18,000円に割り引きます。参加申込書に紹介者(会員)のお名前をご記入ください。情報知識学会会員は参加費が半額となります。

~~~~~

論文募集のお知らせ

2002年12月末発行予定の第4号では、「XMLを用いた情報知識システム」を特集テーマとして、論文を募集します。

XML(eXtensible Markup Language)は様々な分野で用いられるようになってきました。そこで、XMLを用いた情報または知識の表現、処理のシステムなどに関する論文を募集します。皆様奮ってご投稿ください。投稿された論文は査読審査を経て編集委員会で採否を決定します。その手続きは通常と同じです。

論文送付先： 情報知識学会事務局
「特集論文」であることを明記願います。

論文応募締切： 9月末日

採否の通知： 11月末

刊行予定： 12月末

なお、第4号では特集テーマ以外の論文の投稿も通常どおり受け付け、査読審査を経て掲載されます。

年会費納入について

1. ご自分が納入した年月日の確認をしてください。

お手元に郵送された情報知識学会誌の封筒に貼ってある宛名ラベルをご覧ください。最下行に納入年（西暦下2桁）、月（2桁）、日（2桁）の計6桁が印字してあります。印字には、納入日から10日ほどかかります。

2. 「未納」と印字してある場合。

次のいずれかの方法で納入してください。1年分の年会費は、正会員8千円、学生会員4千円です。

- 郵便振込口座 00150-8-706543 情報知識学会（代表 藤原鎮男）
- UFJ銀行 秋葉原駅前支店 普通預金 3586133 情報知識学会（会長藤原鎮男）

所属団体名で支払う場合、郵便振込票の通信欄に個人名を明記してください。銀行振込では個人名が記載されませんので、電子メール・FAX・葉書など、別ルートで団体名と会員名（個人名）をお知らせ願います。

3. 年会費の納入期限

毎年5月末までに納入してください。会計年度は4月1日より翌年3月末までです。退会するかたは新年度に入る前（3月末まで）に電子メール、FAX、葉書などの文書で、退会届をご提出ください。その際、年会費の滞納分はお支払い頂きます。

入会勧誘について

情報知識学会は大勢の方々の入会を歓迎します。どうぞ、お知り合いやお近くのかたをご紹介ください。郵便番号・住所・氏名を事務局へお知らせくだされば、入会申込書・案内パンフレット・学会誌またはニューズレターの最新号を郵送します。あるいは、ご紹介者から直接手渡しして頂いても結構です。

1. 情報知識学会会員の紹介で入会するかたには、入会金を免除します。
2. 入会者をご紹介くださった会員は、年会費が紹介1件につき4千円減額されます。ただし、平成13年4月から平成15年3月までの入会に限ります。

以上、ご不明の点は、ご遠慮無く下記へお問い合わせください。

情報知識学会事務局

TEL: 03-3835-5692 FAX: 03-3837-0368

E-mail: LDE01013@nifty.com

投稿規程

情報知識学会では現在、編集委員会において投稿規程の改定を検討しています。詳細については原稿執筆の際、情報知識学会事務局（後述の原稿送付先）へお問い合わせください。

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）でお出しください。使用したソフトウェアとバージョンおよび著者名をフロッピーディスク表面に記載してください。

論文は、既刊の情報知識学会誌を参考にし、できるだけ以下の形で提出してください。すなわち、冒頭の標題部分と抄録は一段組、本文以下は文中に図・表を入れ二段組とします。技術的に困難であれば、すべて一段組、図・表は別としても結構です。その場合は図・表の挿入位置を本文印字出力に手書きで指定してください。

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。

〒 110-8560 東京都台東区台東 1-5-1 凸版印刷 (株) 内 情報知識学会事務局
TEL: 03(3835)5692, FAX: 03(3837)0368
E-mail: LDE01013@nifty.ne.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 (情報知識学会 事務局) 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。

会員氏名一覧表について

平成15年度は役員改選期にあたります。来年5月の通常総会で承認を受けるために、今秋、会員の皆様から役員候補の推薦を募り、選挙手続きを進める予定です。そのため情報知識学会会員全員の氏名を公表する必要があり、次号情報知識学会誌12巻3号(2002.10.22発行)に会員氏名一覧表を掲載します。

氏名公表を希望されないかたは、ご遠慮なく、本年9月末までに情報知識学会事務局へご連絡ください。その場合は役員候補として推薦を受けることができませんので、ご承知ください。

情報知識学会事務局

TEL: 03-3835-5692 FAX: 03-3837-0368

E-mail: LDE01013@nifty.com

~~~~~

### 情報知識学会役員(平成13~14年度)

|     |         |                 |       |             |
|-----|---------|-----------------|-------|-------------|
| 会長  | 藤原 鎮男   | 東京大学 名誉教授       |       |             |
| 副会長 | 藤原 讓    | 工業所有権総合情報館 理事長  |       |             |
|     | 細野 公男   | 慶應義塾大学 教授       |       |             |
|     | 根岸 正光   | 国立情報学研究所 教授     |       |             |
| 理事  | 有川 節夫   | 九州大学 教授         | 石川 徹也 | 図書館情報大学 教授  |
|     | 石塚 英弘   | 図書館情報大学 教授      | 岩田 修一 | 東京大学 教授     |
|     | 岩淵 幸雄   | 日本創造学会 理事       | 太田 泰弘 | 文教大学 非常勤講師  |
|     | 神立 孝一   | 創価大学 教授         |       |             |
|     | 菊田 昌弘   | (株)シナジー・インキュベート | 代表取締役 |             |
|     | 国沢 隆    | 東京理科大学 講師       | 小山 照夫 | 国立情報学研究所 教授 |
|     | 後藤 智範   | 神奈川大学 教授        | 菅原 秀明 | 国立遺伝学研究所 教授 |
|     | 高橋 靖明   | 凸版印刷(株) センター長   | 田畑 孝一 | 図書館情報大学 副学長 |
|     | 中川 優    | 和歌山大学 教授        | 長瀬 真理 | 静岡大学 教授     |
|     | 名和 小太郎  | 国際大学 客員教授       | 西脇 二一 | 奈良大学 教授     |
|     | 平田 周    | IT 経営研究所 代表     | 福室 淳宏 | 大日本印刷(株) 部長 |
|     | 松田 純一   | (株)日立製作所 主任研究員  | 安永 尚志 | 国文学研究資料館 教授 |
| 監事  | 宍戸 駿太郎  | 筑波大学 名誉教授       |       |             |
|     | 月見里 禮次郎 | 多元空間えんがわ 代表     |       |             |

(五十音順)

### 情報知識学会 評議員(平成13~14年度)

|       |                  |        |               |
|-------|------------------|--------|---------------|
| 芦沢 実  | (株)日立製作所 主任研究員   | 高橋 仁一  | 大日本印刷(株) 開発室長 |
| 田隅 三生 | 埼玉大学 教授          | 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 | 深見 拓史  | 廣濟堂 開発営業本部長   |
| 山本 昌弘 | 法政大学 教授          |        |               |

(五十音順)

---

---

情報知識学会誌 編集委員会

編集委員長 藤原 鎮男 東京大学名誉教授

編集委員

|       |                |        |                |
|-------|----------------|--------|----------------|
| 藤原 讓  | 工業所有権総合情報館     | 細野 公男  | 慶応義塾大学文学部      |
| 根岸 正光 | 国立情報学研究所       | 石塚 英弘  | 図書館情報大学図書館情報学部 |
| 西脇 二一 | 奈良大学社会学部       | 安永 尚志  | 国文学研究資料館       |
| 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 |
| 宇陀 則彦 | 図書館情報大学図書館情報学部 | 神立 孝一  | 創価大学 経済学部      |
| 国沢 隆  | 東京理科大学理工学部     | 後藤 智範  | 神奈川大学理工学部      |
| 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   | 名和 小太郎 | 国際大学           |

---

---

■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。

〒107-0052 東京都港区赤坂 9-6-41 乃木坂ビル 学術著作権協会  
TEL: 03-3475-5618 FAX: 03-3475-5619 E-mail: naka-atsu@muj.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: 978-750-8400 FAX: 978-750-4744 URL: <http://www.copyright.com/>

情報知識学会誌 Vol.12, No.2 2002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://www.jsik.jp/>

---

---

# Journal of Japan Society of Information and Knowledge

## ~~~~~ Contents ~~~~~

### Special Issue: Information and Knowledge

#### Foreward

Special Issue: Information and Knowledge ..... Shizuo Fujiwara... 1

#### Research Paper

SOME ASPECTS OF THE IMPACT OF THE INTERNET ON ACCESS TO INFORMATION ..... Michael W. Hill... 3

DATA, INFORMATION AND KNOWLEDGE (DIK): Challenges in a Changing Society ..... Jacques-Emile Dubois... 10

Performance measures for ranked output retrieval systems ..... Yoshihiro Sagara... 22

A knowledge Management System for the Questions & Answers at the Reference Counters of Libraries  
..... Mariko Iwasawa, Azumi Yoshida, Shoji Ohshita, Yoko Nakanishi... 37

Bibliometric Analysis on Japanese Advertising Activities after the End of the World War II Part I: Phenomenological features extracted from one specialized journal on advertisement ..... Hideya Azuma, Junichi Sato, Shuichi Iwata... 49

The J-STAGE SGML Editing Tool as an Electronization Support System of Academic Journals ..... Osamu Nakagawa, Niichi Nishiwaki... 71

Automatic Extraction of Hierarchical Relationships for Structuralization of Conceptual Structures  
..... Takayuki Morimoto, Naoki Asakawa, Tomonori Gotoh, Yuzuru Fujiwara... 80

#### Opinion

The Positioning of *Johogaku* in Sciences ..... Shu Hirata... 88

Publishing Contents ..... Takushi Fukami... 95

#### Information

The technical terms ..... Kimio Hosono... 98

Report from CODATA ..... Shuichi Iwata... 99

#### Announcement of JSIK

Notices ..... 100

Information for Authors ..... 103

On the list of an association ..... 105

~~~~~

情報知識学会誌 第12巻2号 2002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