

ドキュメンテーションと情報処理*

中村幸雄**

1. 緒言

情報処理という言葉は、広くも狭くも、いろいろに解釈されますが、その範囲のとりようによって、この講演のもうひとつの題目であるドキュメンテーションという言葉との異同がいろいろ考えられます。

しかし現実にドキュメンテーションも、当学会で考えている情報処理も、互いに関係は深いが、やはり一応は別々な分野として国際的にも認められております。しかしドキュメンテーションとはなにか、という点で会員諸氏が十分に理解しておられるかどうかかわかりませんので、情報とか、情報処理とかいう言葉の内容を知らべながら、ドキュメンテーションということの意味を説明しようと思います。

2. 基本的な概念

ドキュメンテーションと情報処理との異同という点については必ずしも定見がないようだし、ことに難しい問題は両方で使う用語に似たものが多くて、抽象的な議論では同じに見えるが具体的には大変違うということが多いわけです。この点は、情報処理とドキュメンテーションで使う言葉の間だけではなくて、ドキュメンテーションと図書館のいろいろな技術を代表する図書館学との間にも似たケースがあります。しかし、やることは具体的によく考えてみると違うと思います。

そういう意味で、一番右の方の隅に図書館学といわれるものがあり一番左の方の隅に当学会でいわれる意味での情報処理がある。その中間にドキュメンテーションがあるというふうな方も許せるだろうと思います。一番の問題は、情報処理という言葉が、ここで皆さん方がお使いになると少しく違った意味で使われる場合が現実に世界にあるという点であります。それからまたドキュメンテーションのほとんど同意語として情報処理という言葉を使う人がアメリカに多く、アメリカでは、ドキュメンテーションという言葉がそれほど

は使わない。むしろこの言葉はヨーロッパの方で発展した言葉でしょう。

情報という言葉の意味の広さがどの辺までなのか私にもよくわかりませぬけれども、とにかく何らかの意味において、我々に役に立つようなものは、みんな情報なんだという風に解釈しますと、ドキュメンテーションでいう情報も情報ですし、この会で普通、情報と皆さんがいわれるものも、もちろん情報であります。ただし、この場合に少し範囲の狭い広いという違いがあると思います。要するにこちらの学会等で主に情報といわれる時には、何かしら、これが物理的な、しかも非常に単純化された存在であるということが暗々の中に仮定されていると思います。ところがドキュメンテーションでいう情報は一般にいったってもっと複雑で、かつ簡単に交換しにくいもの、つまり言語の段階そのものにおける情報というようなことを主に問題にしています。もちろん言語というものも適当な方法をとると全て電気の方でいう意味での信号に転換できるが、転換されて非常に扱い易くなったものを信号とし、コーディングをいかにしたら良いかということがまだわかっていないような状態における対象物、それをドキュメンテーションでは普通、情報とよびます。

だから、つかみどころのないような情報というものありうる。なぜつかみどころがないかというところ、多くの場合、この情報が言葉の形で書かれているからであろうと思います。もちろんドキュメンテーションでいう情報は言葉とは限らず、たとえばパターンであってもよろしい訳です。これを適当なコーディングの方法とか、あるいは Translation によって信号に直しますと、この学会でいう情報処理機械にかかり易い状態になる。

たとえば一冊の雑誌の中に一つの記事があるとすれば、その記事全体はひとつの情報といえますが、一体これをどういう風な形で、信号に直していったら良いかということをやまず考えねばなりません。書いてある文字を一字一字全部追って、適当なコーディングの方法で全部信号に直すことは、もちろんできますが、そういうことをやるのは一般に非常に能率が悪いわけですから、何かしらず言語の段階でこれを簡単にする必

* Documentation and Information Processing, by Yukio Nakamura (Electrical Communication Laboratory, Tokyo)

** 日本電信電話公社 電気通信研究所

要があり、なにがこの情報の中で一番エッセンシャルであるかを見分けなければならない。そうすると当然その中で意味の問題が出てくる。そこでドクメンテーションにおける情報という場合には、言葉としての意味という問題が大きな役割を占める場合がある。たとえば5ページのある論文をよんで日本語の400字の抄録を作ってくれという要求を、たとえば、編集委員会が出す。そうした場合にいかにしたら良い抄録を作れるかということは恐らく、この場合では直接的な意味で、皆さんは情報の処理とお考えにならないかも知れません。しかし、これはドクメンテーションの方では、重要な情報処理の一つの手段であります。したがって「処理」の仕方につきましても、いろんな意味の違いがあるということが、まずいえるのではないかと思います。

そういうような莫然とした話しをまず最初にしておきまして、次に一体ドクメンテーションという概念がどういうものなのか、話した方が良いと思います。それで語原的にいうと、Documentation という言葉*は document という言葉から来ていて“document”というのは普通の意味では要するに“文書”という意味に解釈することが多い。それは別にお役所でいうような概念ではなく、ある意味では資料とよぶものに近いと思います。そしてドクメンテーションでいう意味では、それがやはり物理的な存在に何かしらなっていないければ困るのです。例えば私の頭の中にいろんな情報がたくさん入っておりますけれども、それだけではドクメンテーションでは情報としては扱えない訳です。それは言語の場合でも音として皆さんのお耳に入るとか、あるいは、適当な文字として目に入りうる形になっている、あるいは磁気テープに録音されても良い、つまり人間のもっている五感によって普通に理解しうるような形になった「情報」を前提条件にします。ですから全知全能の人がいても、その人が何も口をきかない、身振りもしないというのでは情報が発生したとはいえない。こういう一般的な意味での「情報」は「資料」と同じだといって、さしつかえないと思います。

次にこういう資料を処理して、扱いやすい形にする。多くの場合、それを蓄積しないと困るので、これを蓄積する。次に蓄積するだけでなく必要な時に必要な観

点から自由に引き出せなければいけない。これは、いわゆるデータ・プロセッシングの一種類だと考えてももちろんさしつかえないが、そういう物が例えば、テープの上の信号になっていなくても良いわけです。すべて文字で書かれた物であってもかまいません。

その場合に扱すのに文字の個々の字づらには別にとらわれずに、多くの場合意味ということを基準にして考えます。これが文章あるいは資料の処理だということにしますとまた別な誤解が起ってきます。たとえば図書館学の方にいわせると図書整理ということと同じではないか、なるほど非常に抽象的ないい方をする。と図書整理ということは「ある情報をもった印刷物を整理して必要に応じて使えるようにしておくこと」という風ないい方でもすればここにいう情報処理と同じになってしまいます。

また、もう一つ非常に似たものとして「Archive」という言葉があります。日本にはあまり archive というものはないので人もあまり気にしないのですが、普通、ヨーロッパなんかで主に archive といいますと、古い文書をとっておくところという風な意味に解釈します。日本語で古文書館という訳をする人もありますがこれは「古文書」である必要は何もなく、この場合は、文書そのものを保管しておくところに非常に重点があり、これを如何にその内容に応じて使うかということ Archive という立場からすると関心がありません。ですから非常にスタティックな考え方です。一方ドクメンテーションは、単に物を取っておくという意味でなくて、積極的に使いたいからそれを整理してとっておくんだという考えであります。その辺のところが旧来の図書館技術との大きな違いであります。もう一つ違いますのはその扱う情報の単位であります。図書館では普通、本といわれるような単位の情報を考えていますがドクメンテーションでは、もっと細かくてたとえば、記事とか論文とか、あるいはもっと細かく個々のデータというようなものが情報の単位であると考えております。

さてドクメンテーションという言葉の普通いわれている定義を述べましょう。それには上に示した意味での情報という言葉をまず仮定します。そうして、この情報について三つの段階すなわち、第1は情報を収集する段階、第2が収集したものを整理しておく段階（整理という言葉にはいろんな意味がありますが一応常識的な意味で考えております）、第3番目には情報

* 英語では documentation, ドイツ語では Dokumentation, ロシア語では dokumentacija といい、ほとんどすべての国で同じ系統の表現をとっています。日本語ではドクメンテーションと呼ぶことにしています。

を今度は人に向けて与える立場、発表、配布する、そういうような段階、この三つを総合したものと定義します。そういう総合した情報の取扱い方がドクメンテーションである、と、考えて良いのではないかと思います。

中間の整理という段階のところにおいて皆さん方が特に普通いっておられるインフォメーション・プロセッシングということが最も密接な関係をもってきます。ところで情報を配布するという問題は具体的には、学会雑誌を作って会員に配ることも一つのプロセスですし、あるいは、そういうことの補助的な手段として雑誌を直接配らずに行う情報配布の手段がありますが、とにかくどうしたら最も能率が良いかということを考えるのは第3の段階であります。こういうことは情報処理学会で考える情報処理とは余りお考えにならないだろうと思います。しかしドクメンテーションの方では、一つの大切なことだと考えます。学術雑誌を作る時に最低どんな条件を満足させなければ情報の流通がうまく行かないかということを考え、例えばこれを規格にします。そういう規格をいかにして作ったら良いかとか、どういう内容を持たせるべきだとか、あるいは、そういう規格間の調整ということをどうすべきかというようなことは、ドクメンテーションにとって重要な一つの分野に属します。

こういうことは情報処理学会の情報処理には余り関係しないかも知れないけれど、一般に学会員としての皆さん方には相当関心のあることであります。たとえば論文の発表形式として雑誌に投稿し、印刷し、会員に配るという形式は、本当は能率が悪い方法なのです。もっとそれを能率よくするためにはどうしたら良いか、現在世界中でいろいろ論じられておまして、幾つかの、案が出ておりこれを国際的な規模で実行するには、どうしたら良いかが、しきりに議論されております。

ドクメンテーションにおいては言葉の意味が非常に重要なので、その意味ということとは当然概念ということに結びつきますし、その概念が現実言葉でどういう風に表示されるかという問題にも結びつきます。例えば「犬」という言葉自身がドクメンテーションの直接の対象には実はならないのです。それよりは、「犬」というものが含んでいる概念を扱い、「犬」と漢字で書いてあっても、ローマ字で「Inu」と書いてあっても英語で「dog」と書いてあっても、フランス語で

「chien」と書いてあっても、かまわない。あるいはまた、それが適当な分類法に従って、たとえば 599.742.1 という分類の一種のコードであってもかまわない。我々は、それらを見て、みんな共通な「犬」という概念を考えるわけです。これらの間には一定の対応関係がありますが、対応関係自身には、それほど興味を持っていない。ですからコードとコードの間の変換は具体的には、入用なんですけれども、抽象的な意味でドクメンテーションを考えるときには、ほとんど問題になりません。したがって、分類法を考える時には、言語にならない分類法を常に考えます。国際的に共通な概念というものを対象にするわけなのです。

3. 分類

さて分類ということとはドクメンテーションの中で非常に重要な地位を占めるものであります。分類ということの意味も非常に広いので、たとえば雑誌「情報処理」の本年の Vol. 2, No. 2~3 を見ますと、淵さんの書かれた分類という論文が出ております。この論文の中では分類ということとは、要するに不規則に並んでいるものをある一つの目安に従って規則的に並べ換えることである、リアレンジメントが分類なのだと考えておいでになります。ドクメンテーションでは、それは分類の一部分にしか過ぎないと考えます。要するに、いろいろな概念がわれわれの前にあるとき、個々の概念がわれわれの知識とか知恵の全体の中で、どういうところに位置づけたいかということを考えることが分類だというふうに考えます。したがって、これはまだ概念を取扱っている段階であります。しかし抽象的な概念といっても、具体的には言語で表示されなければならない。したがって当然のこととして、概念の言語による表現の問題（言語の段階）がつきまってくる。次に現実の言語だけでは不自由な面がある。そこでそれをもっと便利な記号によって表示して扱いやすいようにしてやる。これが分類の「記号の段階」といわれるものであります。

たとえば犬というものを考えるとき、動物学における分類の例をとりますと（第1表参照）、生物という

第1表

| | | | | | | | | | | |
|----|---|-----|---|--------|---|---------|---|-----------|---|----|
| 生物 | > | 動物 | > | 哺乳動物 | > | 食肉目 | > | 裂脚亜目 | > | 犬科 |
| 59 | > | 599 | > | 599.74 | > | 599.742 | > | 599.742.1 | | |

大きな枠があって、その中に動物、またその中に哺乳

動物がある。それから食肉類とか食肉目とかいうものがある。さらにその下に(いろいろやり方があるんですが)裂脚類、さらにそれを分けて犬科、そういうふうに概念の上で分類の操作を加えていく。それをまた記号の段階で表示するときにはちょうどこれと同じように、細かくなればなるほど、表示する記号の桁がふえてくる方式(あるいはその逆であっても、もちろんかまいません)をとることにします。第1表に例があがっているように、概念が細かくなるに従って表示する記号を長くする仕方により 599.742.1 という動物の分類上のある地位がここで示されます。しかし同時に犬という言葉にはまだいろんな見方があり、たとえば家畜としての犬というものがある。これは動物の分類形における犬に近いが、動物一般とは別の観点がある。そういうものを同じ分類体系では 637.7 で表わします。

また犬という言葉の中には、たとえば「あいつは警察の犬だ」というような使い方もある。これは言葉としては同じですが概念においては違うので 327.84 というところへ入る。

このように一つのものを見る見方というものはいろんな観点があって、分類というものは決して一通りだけのものではないということが、根本的な前提になります。ところがそういう認識は割合に新しいもので過去の図書館の技術にはあまりなかったわけです。

また、分類だけが必ずしも唯一の整理の手段ではなく、よく使われる別の手段に索引があります。普通見受けられる形式のものは大まかな概念を扱うものですが、もっと細かな概念を活用するために、索引の方法は最近いろんな進歩を遂げております。

複雑な内容をもった文章をいかにして符号化して、索引するかという問題は一種の意義の分析というものを伴ってきます。たとえば「猫が犬をかむ」という文章があるとすれば、この中から意味のある(significant)概念として、たとえば、犬と猫と「かむ」という動作の三つをとりあげます。しかし「猫」と「犬」と「かむ」を並べただけの索引では、それは猫が犬をかんでいるのか犬が猫をかんでるのかわからないわけです。そこで何かもう少し高級な索引形式をとろうと思つと、おのおの言葉の文法的な位置、あるいは単語の役割というものを考えなければならぬ。そうした場合に第2表のような並べ方をして「猫」が主体であり「犬」は客体であるという何か符号をつける。「かむ」ということは動作を表わしているのだという符号

第2表

| 概念 (ねこ) | 並列式索引 |
|------------|----------|
| (かむ) | ね こ |
| (いぬ) | い ぬ |
| 意味を加味した索引法 | か む |
| ④ | ⑤ |
| ね こ (主体) | い ぬ (主体) |
| い ぬ (客体) | ね こ (客体) |
| か む (動作) | か む (動作) |

をつけると、明かに猫が犬をかんでいるのだと表わせます。

もう少し複雑な例(第3表参照)を考えてみます。

第3表

| | | |
|------------|----------------|-----------------|
| 温度計 | — | 装置, 測定, 熱 |
| M | CH | 装置, 機械 |
| M | SR | 測定 |
| R | HT | 熱 |
| Infix code | A | その一種, Uが作る, を作る |
| | W | によって作用をうける |
| 上以上を総合して | MACH MUSR RWHT | .004 X .001 |
| 電圧計 | MACH MUSR LWCT | .006 X .001 |

これは Western Reserve 大学 (WRU) で行なっている情報のプロセッシングの一つの例であります。ここまでできますと非常に皆さん方のおっしゃる意味の「情報処理」に近くなってきます。これはある複雑な文章で書かれたものの内容を、その文章全体からも引出せるし、あるいはその中の任意の重要な概念からも引出せるようにしようという考え方です。具体的には実際にデータ・プロセッシング・マシンを使っておりますが、複雑な内容をもった文章を、どういふふうにコード化するかが大きな問題になってきます。

まず単語を符号化する。次に単語と単語ででき上っている文章の符号化をするという、二つの段階に分けて考えます。単語の一番簡単な符号化は温度計という日本語ならその発音を適当なコーディングの方法で直す。たとえば六単位の電信符号を使う。もしそうすれば「温度計」と「寒暖計」という言葉は別々な信号になり、その間に全然関係がつかないわけです。WRUの方法では、その関係が信号の上で、はっきり出た方がよらしいということを考えて、温度計というものを次のようにコーディングします。

温度計というものは大きくみると「装置」である。しかも「測定する」装置である。そして「測定の対象」は「熱」である。このおのおのの概念にコードをつけます。装置とか機械とかいうものは M CH と書く。測

定ということは MSR と書き、熱というものは RHT と書く。ところがこれだけの記号を三組並べたのでは意味の上でどう結びついているかわからない。そこで上のコードの中であけてあるところにそういう意味の関係を表わす別のコードを入れます。これを Infix code と呼んでおり、たとえば A という字はその示された概念の内の一種類を表わす場合、U という字はその概念が作り出すもの、あるいは概念を作るようなものを表わす、W という字はそれによって作用を受けるものを表わすなどときめます。つまり現実の言語でいいますと、たとえば前置詞というようなものに当るわけでありませう。そういうコードをはめると、温度計というようなものは MACH, MUSR, RWHT となります。そういう機械はいろいろあるので、個体を表わすために、さらに 004 X .001 というような記号をつけます。004 X は、熱の形態の一種類として温度というものを考えるということを表わしている。温度を測る機械というものは他にいろいろあるので、寒暖計と普通われわれが知っているものはその内の第一のものであるという意味で .001 がついたわけでは

同じような考え方で電圧計を考えると前の二つのコードは当然同じになるが測定する対象が電圧であるということで、三つ目のグループから先が変るはずで、明かに単語そのものを扱っているのではなくて、その単語の奥にある概念というものをコード化しようという考え方であります。

次にそのような単語(あるいは単語の表わすコード)がいくつか集まって文章を作るわけですが、これもただ並べただけでは文法的な関係がわかりませんから、何か別なコードをまた入れようじゃないかと考えるわけでは

第 4 表

| |
|---|
| -KEJ (合金), KUJ (鉄), KUJ (ニッケル), -KAM (冷却) |
| KEJ 処理加工を受ける材料 |
| KUJ 成分 |
| KAM 過程 |
| 意味: 鉄(と)ニッケル(とを成分とする)合金を冷却する。 |

こでくってありますが、この意味はここに合金ということを表わす適当な単語のコードがくることを示しています。こういう単語を今度は組み合わせるのですが、ここでは「鉄・ニッケル合金を冷却する」という例があげてあります。鉄・ニッケル合金を冷却するというのをもう少しパラフレイズしますと、それは「鉄とニ

ッケルを成分とする合金を冷却する」といえるわけでは、そこで合金ということのコードの前に KEJ というコードがありますが、これは処理加工を受ける材料、つまり合金という操作を受ける材料を表わします。

KUJ というのは成分ということを表わすコードであります。したがって成分である鉄、成分であるニッケル、これによって合金という処理を受けて作った合金、それが鉄、ニッケル合金であります。次の KAM というのは過程、プロセスを示します。これらを組み合わせて、鉄、ニッケル合金を冷却するというようなことを表わします。

このような方法でコード化する元の文章は、非常に丁寧なものであってもかまわないし、雑な書き方であっても、同じくこの程度の内容のコードになってしまうわけでは、そういうふうでコード化したものを蓄積しておけば、どれか一つのコードだけでも引出すことができますし、コードの組み合わせとして引出すこともできる。そういうようにしていくと、こういう情報を含んだ文献とか記事とかがどこにあるかということ、適当な所在のコードを添えておきますから引出せるわけでは、ここまでいきますと、これはまさにデータ・プロセッシングになります。しかしそれがすべてでなく、ドクメンテーションとしては、この操作はむしろおしまいの方の段階なのであって、その前に、いかにしたらば何かわけがわからない非常に複雑な内容をもっている情報を分析して要素的な言葉に分けて、その言葉を適当な概念に純化して、その純粋な概念の組み合わせとして情報を引出すにはどうやったらいいか、を考えるのです。ですからデータ・プロセッシングの機械には必ずしもかけなくてもいい、たとえば目で見て文字で書かれた記号だけで、そういう操作を行なうことも可能であります。現実にはむしろそういうシステムの方が現在では多く行なわれています。そういう肉眼で扱うシステムといまのようなデータ・プロセッシングの機械にかけるものと、どういう場合にどっちが得になるかということも考えることも、ドクメンテーションの中では非常に大事な仕事であります。

この WRU の方法のように、磁気テープを使うプロセッシング・マシーンというのは、日本では試作をなさったところが現在二箇所しかない。アメリカでもそういうものはそう多くはなく、むしろ IBM のカード式の装置を相当使っている。しかしパンチカードによるとしても、IBM カードが唯一のものではない。もっ

と簡単で手操作は耐えるハンド・ソート・カードで、情報の処理が相当広範囲にできます。このカードといまの IBM のカード式機械の現在の機能と能率を前提にした比較検査も行なわれております。非常に簡単な例をとると、情報の数が1万以下のような場合には IBM の機械はそんなに能率がよくない。たとえばスピードからいうと、ハンド・ソート・カードに比べてせいぜい2倍から3倍ぐらいのスピードしかない。したがって人件費や設備費を考えると、必ずしも IBM の機械を使うことは有利でないという判断も出てくるわけです。実際にドキュメンテーションを業務として扱う場合には、このことは非常に重要になるので実証的な研究も行なわれております。

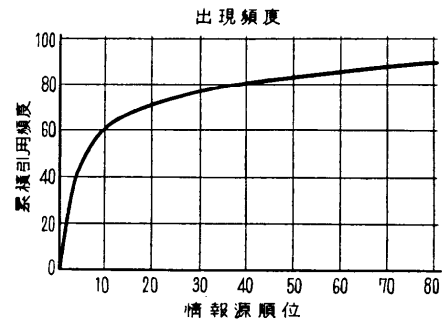
4. 情報処理の性質

次にもう少し違う性質のことを申し上げようと思います。特定の人、機関にとって情報源はなんでもかんでもみんな等しく価値をもっているということはないわけでありまして。たとえば雑誌「情報処理」は小説家に対してはほとんど値うちがないでしょう。ところが私にとってはたとえば、アリストテレスの哲学というものはあまり情報の価値がない。そういう相対的な違いというものはあるわけです。そこで特定の人、あるいは特定の分野に働く人、またはそういう人の団体を考えますとき、個々の情報源はどんな有用さをもっているかということ、資料をどのくらい自分のところで持っておいたらいいか、ということの判定する上に重要な要素になります。そういうことの具体的な研究が行なわれています。第1図～第2図はエレクトロニクスの分野において電気通信研究所でとった統計の結果であります。個々の雑誌が使われる度合というものを適当な尺度で表わしその順位に並べてみます。

「使われる度合」というものを表示するいろいろな方法がありますがけれども、ここでやりましたのはエレクトロニクスの分野における論文の著者が行なった引用の統計をとる方法であります。引用されたものは有用な情報であったと一応考えていいわけですから、引用された度数というものを有用度の一つの代表だと考えることができます。そこである部門でいろいろな雑誌をずっと見まして、全部見るわけにはいきませんから、適当なサンプリングをして論文のおしまいに出てくる引用文献について、それがどんなソースから取られているか、また時間的にみていつごろのものがたく

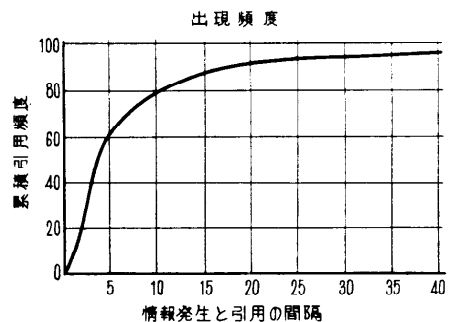
さん出てくるかということ調べたわけなんです。

第1図は雑誌を情報源順位に並べて、何番目の雑誌までで、全体のうち何パーセントの情報が得られるか



第1図

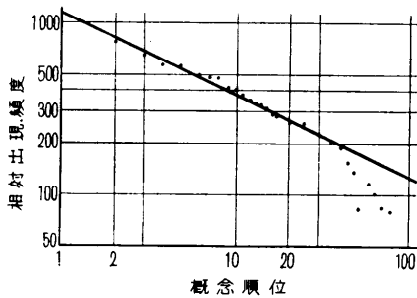
という累積引用頻度を描いたものであります。この結果を見ますと、たとえばわれわれが十種だけの雑誌を持っていれば、必要な情報の60%はまかなえることがわかります。100%を希望しますと、非常にたくさんの雑誌がいる。これは当然であります。たとえば経済的にある資料室とか図書室でも設計しようと思うときには、どのくらいの雑誌を常時持っていたらいいかということ design するときにはこれが大切なデータになるわけです。たとえば80%をまかなうためには40種類、90%にしますと大体7~80種類あればよい。それから先をまかなおうとすれば非常に数が多くなるので、そんなたくさんの雑誌をめいめいの機関が持つことは、非常にロスである。むしろそういうものは、たとえば科学技術情報センターというような、共通の組織に持ってもらい必要なところだけの複写をもらう方が安いという経済計算ができます。次の第2図



第2図

は、引用してあるものが引用したときから何年前のものであるかを見たわけであり、やはり累積頻度に直して書きますとおもしろいことがいえます。たとえば必要な資料の60%は5年以内のものである。10年までそれを拡げると80%になる。90%を確保したければ20年ぐらいまでのものでたくさんであります。雑誌は大体20年ぐらいから先というものは非常にわずかしか利用しないから捨ててもほとんど害を及ぼさないというような判断ができるわけであり、近代的技术図書館ではそういうことを実行しているわけであり、

次の第3図はある分野で、個々の論文を代表する重



第3図

要な概念を多数の論文について調べて、どういう概念が頻繁に出てくるかという統計をとった例であります。そういうものを見ますとこれを両対数の方眼紙に書きますとほとんど直線の上ののっている。そうしますとさがすときにはある程度どの辺まで重要な概念をひろっておけばあとは無視しても何%の危険率しかおこさないということがいえるわけです。

5. 抄録の問題

抄録の問題は詳しくは申しませんが、これは現実には重要な問題であります。たとえば本会誌「情報処理」では抄録というものが載っていない。これはある基準によりますと、学術雑誌として少し極端にいえば、下の下である、というような判断もあるわけです。しからば抄録が載っていれば、何でもいいかということ、抄録にはやはり抄録の品質というものがあるわけ、その品質を上げるにはどうしたらいいか、というような問題があるわけです。たとえばその抄録の中に「ここに著者が述べるものは、著者が20年来にわたって研究したもので、学術的に価値大なるものがある……」など

とばかり書いてあって、肝心なことは2行ぐらしか書いてないような抄録があるとすれば、これはまた下の下であります。

定量的な判断ができるかどうかということまで実はまだ行っていないのですけれども、質的な判断はできます。質的に高い抄録であるためには、最低どういう条件だけは守るべきだ、という基準が現在ではあるわけです。これもドクメンテーションという意味での情報処理では非常に重要な問題であります。

もう一つ私が先程学会として興味があるといった問題を申し述べます。それは情報の発表あるいは配布の形式あるいは方式という問題であります。現在一番問題になりますのは、情報が現代ではあり過ぎて困っている、われわれは情報の洪水の中に溺れかかっていることです。われわれが10年先に必要になるものはいまからわかっているわけではないんですから、10年先に必要になったときに、適当な手段があって、必要なものが短時間に探し出せ、それがどこにあるかということは別に憶えていなくても、適当なインデックスなり何なりを見て、探し出して注文する。注文するとそれが一定の時間の間にわれわれのところへ届いてくる。そういう組織をもつことの方が重要なわけです。またその手段を具体的に考えることが非常に重要なことであります。

雑誌をどの会もこの会も、われもわれもと出すことは論文を発表する人には非常に楽しいものですが、見る方からいいますと、非常に手間がかかる。また探す方からいいますと、必要なときにはすぐ捜せるし、捜したものについては、われわれの入用な程度に細かい情報が欲しい。取扱う上からいいますと、いろいろなものがたくさん出てはこまるということ、特定のものについては細かい内容まで知りたいということは互いに矛盾する条件です。従来から単に情報の発表ということと情報の入手ということには、それぞれいろんな複雑な機能があるにもかかわらず、両方を無理に一つにしていたということが、この困難を生んでいる一つの理由であろうと思います。そういう意味で、われわれはその情報の発表および配布ということに機能の分化を考えなければいけない。一番簡単な方法は、情報がどこにあるかを知らせることとその情報の内容を自分が見て知るということは別な操作だと割り切ることです。前の方の役にたつものをわれわれは二次情報、後の方の情報そのものは一次情報とっております。

したがって二次情報を見ること、一次情報を見ることは全然別な機能であります。

その別な機能を行なう別々な組織というものがあるていいではないか、そういうことを考えるわけでありす。たとえば具体的な一例として、学会雑誌というものを非常に部厚くして、それにみんなフルペーパーを載せるということは経済的ではない。むしろわれわれはどこに行ったらばフルペーパーがあるということだけ知らしてもらい、入用になったとき適当な手続きをとれば、そのフルペーパーがくる。たとえば IRE は何万部印刷したからといってもその中に載っている記事を何万人の人が読むということは想像できない。特定の記事についてはおそらく数百人ぐらいの人しか読まない。そうしますと、大勢の人に読まれるものだけを学会誌に載せて、それ以外のものは適当なセントラルな機関があるとすれば、そこに元のものを置いておき、必要な人にだけ必要なときに、短かい時間で、そのコピーを送ればいいわけです。もちろんこれは現在そういう話を日本でしますと夢のような話だといって簡単にかたづけられるのですけれども、しかし現実にはそういうことも考えなければならぬ時代にきているわけです。そういう組織としては日本なら、日本科学技術情報センターは、そういうことに向く有力な候補者に違いないけれども、現在のところはいろいろな制約からそういうことが行なわれていない。むしろわれわれは、科学技術情報センターをはげまして、そういうことをするようにし向けなければいけない。そういうような意味で、情報を流通させる手段を広くわれわれが共通の問題として考えるべき時期がきているということを、ここで私は申し上げておきたい。

6. ドクメンテーション機関

こういうドクメンテーションの機能を専門的に扱う人というものは、従来ほとんどなかったわけですが、最近はそのような人が増えつつある。こういう人をドクメンタリストという言葉で表わすとして、そういう人の機能は図書館における司書つまりライブラリアンと似てるとか違うとかいう議論がどこの国にもあります。私はそれはある点ではっきりした違いがあると思います。詳しいことは時間もないので省いておきます。

もう一つの問題は、情報に対する要求は研究の場合特に大きいので、ドクメンテーションということがある意味で、研究の管理ということに密接に結びつきます。しかし、本質的にいえばやはり研究の管理ということとドクメンテーションとは別物であります。

いろんな段階がありますが、ドクメンテーションの非常に高級な段階は判断を伴う業務であります。しかしこの一步手前の状態は要するに情報サービスと普通いわれております。企業体の場合には、情報サービスという非常に重宝なサービスを使って、その上に立って情報の価値判断を行なう。その判断に従って、しからばわれら何をなすべきかということを考える。これがドクメンテーションにおける最高の段階だといってさしつかえないと思います。そういうような意味で、ドクメンテーションということの一つの専門的な職分として扱う団体がいくつか最近はあるわけです。皆さん方に名前だけ覚えていただこうと思いますのは FID* で、本部がオランダにあって、日本では現在のところ日本学術会議の中にある、ドクメンテーション研究連絡委員会がその窓口となっております。

それからドクメンテーションの仕事を現在日本で本当の業務として広範囲に扱っている機関は、日本科学技術情報センターだけではないでしょうか。たとえば国会図書館はそういうことをするんだと称していますが、實際上、国会図書館にそれだけの能力はありませんし、また国会図書館があんまりそんなことにまで手を伸ばすことは、法律上はともかく、国会図書館本来の業務からいって、私は適当ではないと思います。それから、ドクメンテーションを一つの技術として扱う日本ドクメンテーション協会というものがあります。これはつい最近社団法人になって、技術としてのドクメンテーションの振興をはかるために一般の学会と同じような活動をしております。

今日私はどんなお話をしたらよいかわかりませんが、情報処理学会と日本ドクメンテーション協会と両方の会員でありますので、二つの会が目ざす目的の間に、どんな関係があるかを知っていただこうとして、ドクメンテーションというのはおおよそこんなものだという、ばく然としたムードをお伝えしようとして簡単なお話をしたわけでございます。

* Fédération International de Documentation の略