

2 . 保有情報のデジタル化

2 . 1 デジタル化方式の選択

紙媒体のデジタル化には表 2.1.1 の各方式が考えられ、コスト、手間、データ量と共に文字列検索やレイアウト情報の必要性等考慮しながら選択する必要がある。特に画像を拡大したときの平滑性に関しては一般的に以下のことが言える。

方式 はページ全体がビットマップ画像になるので、拡大表示すれば不鮮明になる。

方式 ~ は文書がデジタルベースなので、アウトラインフォント(文字の輪郭を幾つかの点の数値で作成)により拡大表示しても滑らかな輪郭の鮮明な文字になる。

方式 、 の図表もビットマップ画像になるので、拡大表示すれば不鮮明になる。

方式 の場合は、グラフやイラスト等を表計算ソフトや DTP(Desk Top Publishing) 関連ソフトで作成していれば、ベクタグラフィック画像(数値情報に従って画像を作成)により拡大表示しても鮮明である。

ここで、表 1.3.1(a)に対して方式を検討する。

A 及び B の 1)の助成研究テーマ関連は、研究テーマ名や研究者名での検索が可能であればほぼ目的は達せられる。1 テーマ当り 3,4 ページであり、内容的に立ち入って検索する必要性は薄い。今回は最近 2 年間のものを方式 で対処する。

B の 2)の音の散歩路も、写真が多用された特異なケースであり著作権もフリーであることから取り上げた。最近 2 年間のものを方式 で対処する。その他の巻頭言や対談は実験的意義も薄く著作権上の問題もあり今回は対象外とした。

表 1.3.1(b)の報告書は、内容的に検索機能が必須である。そこで財団が独自に取りまとめた No.16,17 は既に文書がデジタル化されていて著作権もフリーなので今回の対象とする。これ以外のものは方式 又は をとる必要があるが、文書をデジタル化する際、以下の問題点があり対象としなかった。

方式	紙媒体の構成		処理内容	特 徴								備 考	
	文書	図表 写真		コスト	手間	データ量	文字列 検索	再利用	レイアウト 情報	文書 拡大の 平滑性	図表 拡大の 平滑性		
	アナログ	アナログ	ページ単位で全体を一度にデジタル化する			×	×				×	×	・補助手段として文書をOCRでデジタル化して併設し、文字列検索に供することもできる
	アナログ	アナログ	文書と図表・写真を分離して各々をデジタル化した後、双方を合成する	×	×							×	・文書をデジタル化するに際し、OCRにしる手入力にしる黙視による確認作業が必要となりミスを犯す危険がある ・文書と図表・写真が入り組んだ複雑なレイアウトの再現は難しい
	デジタル	アナログ	図表・写真をデジタル化した後、文書のページに挿入する									×	・文書と図表・写真が入り組んだ複雑なレイアウトの再現は難しい
	デジタル	デジタル	何もしない										

表 2.1.1 紙媒体のデジタル化方式

文書のデジタル化は、OCRソフトを利用するか、再度手入力によるかのいずれかの作業が必要となる。しかし、OCRの認識率も90～95%が現状³⁾であり、方式の備考に記した様に文字列検索だけに用いるならまだしも、方式の様に報告書の原文として使用すると、誤りを修正して100%の精度を目指さなければならない。結局、誤り修正の手間を考えると、最初から手入力の方が安く上がるケースがほとんどと言う。印刷・出版業界でよく利用される手法は、2人のオペレータによる手入力結果をパソコンで比較して誤りを見つけ出す手法であり、マンパワーを必要とする。⁴⁾方式でも検索精度を上げるにはOCRの認識ミスを少なくする努力は必要であろうし、文書も画像として扱うのでデータ量が多くなり、ページ数の多いものには向かない。

以上の点を考慮して、今回は表2.1.2の様に保有情報のデジタル化を実施する。

<p><u>A. 研究助成事業</u></p> <p>助成研究成果報告概要</p> <ul style="list-style-type: none"> ・ 昭和 59 年度～平成 8 年度 ・ 平成 9,10 年度 ・ 平成 11 年度 以降 	<p>研究者からの要望 コピーサービスで対応</p> <p><u>方式</u> に対応</p> <p>デジタルベースに切換え 将来は方式、 に対応</p>
<p><u>B. 機関誌「サウンド」</u></p> <ul style="list-style-type: none"> ・ 第 1 号～第 12 号 ・ 第 13,14 号(平成 10,11 年 1 月) ・ 第 15 号(平成 12 年 1 月) 以降 	<p>研究者からの要望 残部を送付</p> <p><u>方式</u> に対応(助成研究テーマ関連、音の散歩路)</p> <p>デジタルベースに切換え 将来は方式、 に対応</p>
<p><u>C. 調査研究事業</u></p> <p>表 1.3.1(b)において</p> <ul style="list-style-type: none"> ・ No.1～15 の報告書 ・ No.16 の報告書 ・ No.17 の報告書 	<p>研究者からの要望 残部を頒布</p> <p><u>方式</u> に対応</p> <p><u>方式</u> に対応</p>

表 2.1.2 保有情報のデジタル化の方式選択

他方、バックナンバーに対するアクセスの利便性が電子配信のメリットという捉え方も広くあり、今回の対象外のものも暫時デジタル化していくことが必要であろう。

2.2 検索方式の考察

検索の方式には、フリーキーワードによる検索、階層分類による検索、或いは自然語入力で検索する自動インデクシングによる検索等いろいろあるが、現在の保有情報の規模からフリーキーワードによる検索で十分であろう。規模の拡大とともに他の検索ソフトを付加して機能の向上をはかって行けばよい。