

厚生科学研究費補助金(子ども家庭総合研究事業)

分担研究報告書

母子保健情報の登録・評価に関する研究

「心身障害研究・子ども家庭総合研究報告書」のデータベース化に関する研究

分担研究者:中村敬(日本子ども家庭総合研究所研究企画情報部)

研究協力者:斉藤進(日本子ども家庭総合研究所母子保健研究部)

庄司順一(日本子ども家庭総合研究所研究企画情報部)

恒次欽也(愛知教育大学教育学部)

横田千秋(三井化学株式会社)

モニター:作田亮一(濁協医科大学越谷病院小児科)

瀧川逸朗(東京都立大塚病院小児科)

長坂典子(母子愛育会総合母子保健センター研修部)

上石晶子(東京都母子保健サービスセンター)

松浦賢長(京都教育大学)

【研究要旨】

心身障害研究および厚生科学研究子ども家庭総合研究はわが国の母子医療・保健・福祉に関するトップレベルの研究として、その成果は施策や研究・教育に活用されている。本研究ではこの研究事業の成果を記した研究報告書を永久保存することと、これらの成果を広く普及することを目的として、これらの報告書の電子化およびデータベース化の方法を検討した。研究方法は、すでに印刷された報告書が対象になる過去分の「心身障害研究」報告書と、各研究者から報告書を電子データとして収集可能な平成10年度以降の「子ども家庭総合研究」報告書の2分野に分けて検討し、平成10年度にCD化の方法論を確立した。

今年度は、平成10年度の研究成果を踏まえ、昭和50年～平成9年(23年分)の過去分「心身障害研究」報告書を解体し1ページごとに画像データとして、イメージスキャナで取り込み電子データ化を図った。また、新規の平成10年度の子ども家庭総合研究報告書については、研究者から文字情報を生かせる形式で、直接電子データで収集し、データベース化を図った。

【見出語】

厚生省心身障害研究 厚生科学研究 子ども家庭総合研究 報告書 データベース
CDROM 画像データ acrobat

A. 研究目的

過去27年に及ぶ「心身障害研究」および平成10年度から実施された厚生科学研究「子ども家庭総合研究」の成果が広く有効に活用されるためには、より多くの関係者に報告書

が配布され、研究内容が手軽に入手できる体制を整える必要がある。

このためには、報告書を電子化し、検索機能を備えたデータベースとして構築することが最も有効な方法と考えられる。電子化された報告書は、従来の印刷物に比べ、コンパクトな媒体に大量のデータを収納することが可能であり、また、作成方法を工夫すれば、強力な検索機能を備えることができる。

本研究班に課せられた目的は、過去の心身障害研究報告書をデータベースとして、実用化するとともに、これからの「子ども家庭総合研究」報告書をデータベース化するための技術的方法論を確立し、さらに、その方法に基づいて、実際に報告書のCD化を図る事である。

B. 研究方法

研究の対象は、1)過去の心身障害研究報告書(平成9年度まで)と、2)新規の子ども家庭総合研究報告書(平成10年度以降)の2つであり、前者はすでに製本されている報告書であり、後者は各研究者から電子データとして収集可能な報告書である。

研究の方法は、1)に関しては、平成7~8年度心身障害研究「保健・医療・福祉にかかわる情報と社会資源の有効活用に関する研究」、および平成10年度本研究において、電子化に関する方法論を確立した。2)に関しては、平成10年度本研究において、現在普及している電子化方法と検索ソフトについて検討し、報告書の電子データの収集方法とデータベース化の方法論を

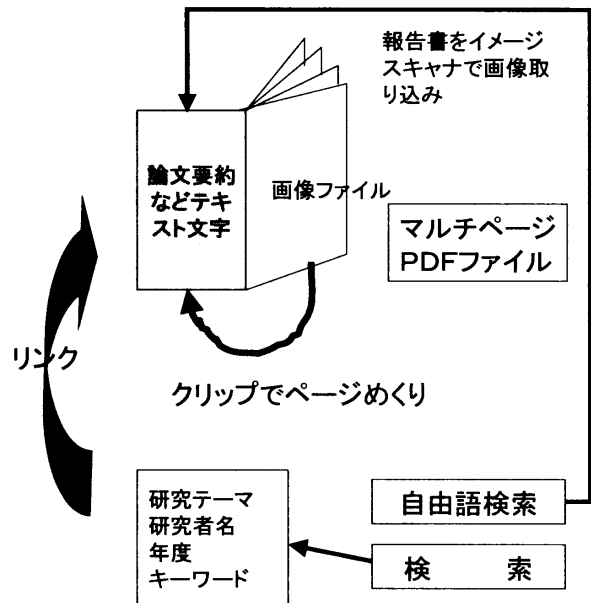
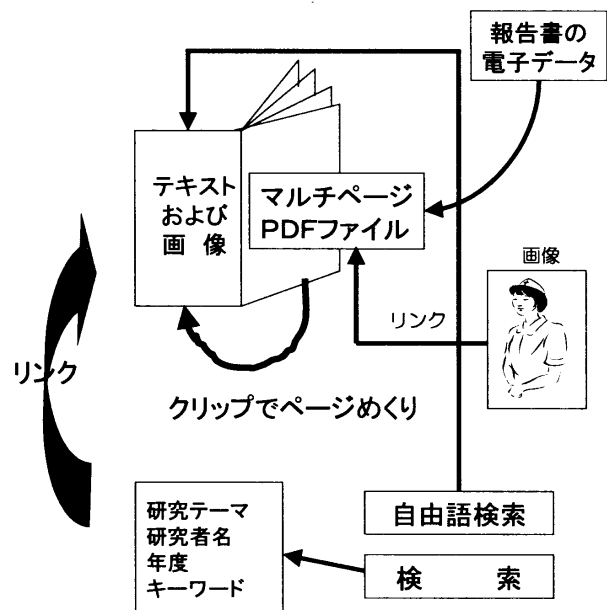


図1：過去分報告書データベース構造



新規分報告書のデータベース構造の基本形
これに、一部画像ファイルで収録したデータも混在する形式。

図2：新規分報告書データベース構造

確立した。

本年度は、平成 10 年度で確立した方法論(図 1、図 2)に基づき、

1)過去の心身障害研究報告書のうち、イメージスキャナで画像として取り込むことが可能な報告書について、1 ページごとにスキャナで取り込み、マルチページ化して電子データ化を図った。昭和 46～49 年の報告書は、製本化し保存しており、長い年月を経て、紙質が劣化し、変色したり、中には手書きの原稿もあり、画像データとして取り込んでも不鮮明で判読不能な個所が多くなると考えられ、電子化の対象から外した。したがって、本年度の作業は、昭和 50 年から平成 9 年までの 23 年分の報告書について、画像ファイルとして電子化を図った。

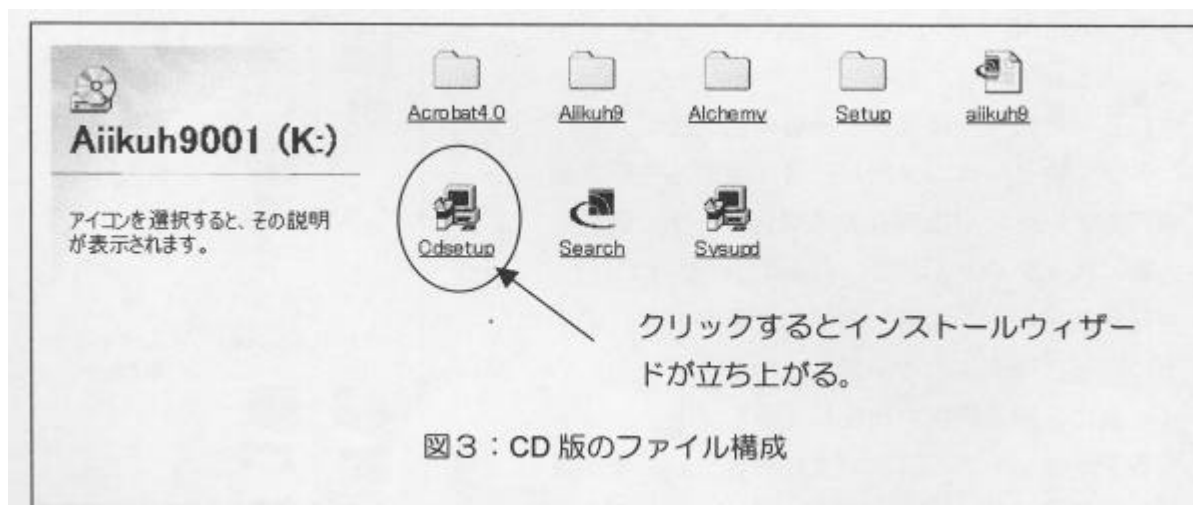
2)新規分報告書(平成 10 年度子ども家庭総合研究報告書)については、報告書と一緒に収集した電子データをもとにデータベースを作成した。

C. 研究結果

本年度作成したデータベースについて、過去の心身障害研究報告書と、新規分子ども家庭総合研究報告書の各データベースに分けて紹介する。

1)データベース検索ソフトのインストールについて(図 3)、

データベース検索ソフトは Alchemy release6 を用いている。対応機種は IBM/PC 互換機、PC-9800 メモリ 16MB 以上(64MB 以上推奨)であり、OS は Windows95、98、NT4.0 のいずれか、マルチセッション対応の CD-ROM ドライブ、ハードディスク空き容量は 25MB 以上を必要とする。また、アクロバット・リーダーをプラグインしておく必要がある。現時点では Macintosh OS には対応していない。



2)データベースの表示

この画面は、今年度作成した過去分心身障害研究報告書の平成 7～9 年度版と平成 10 年度

子ども家庭総合研究報告書のサンプル版を DVDROM にコピーしたものである。したがって、実際にリリースする CD 版 (各年度 1 枚) と若干異なることをお断りしておく。

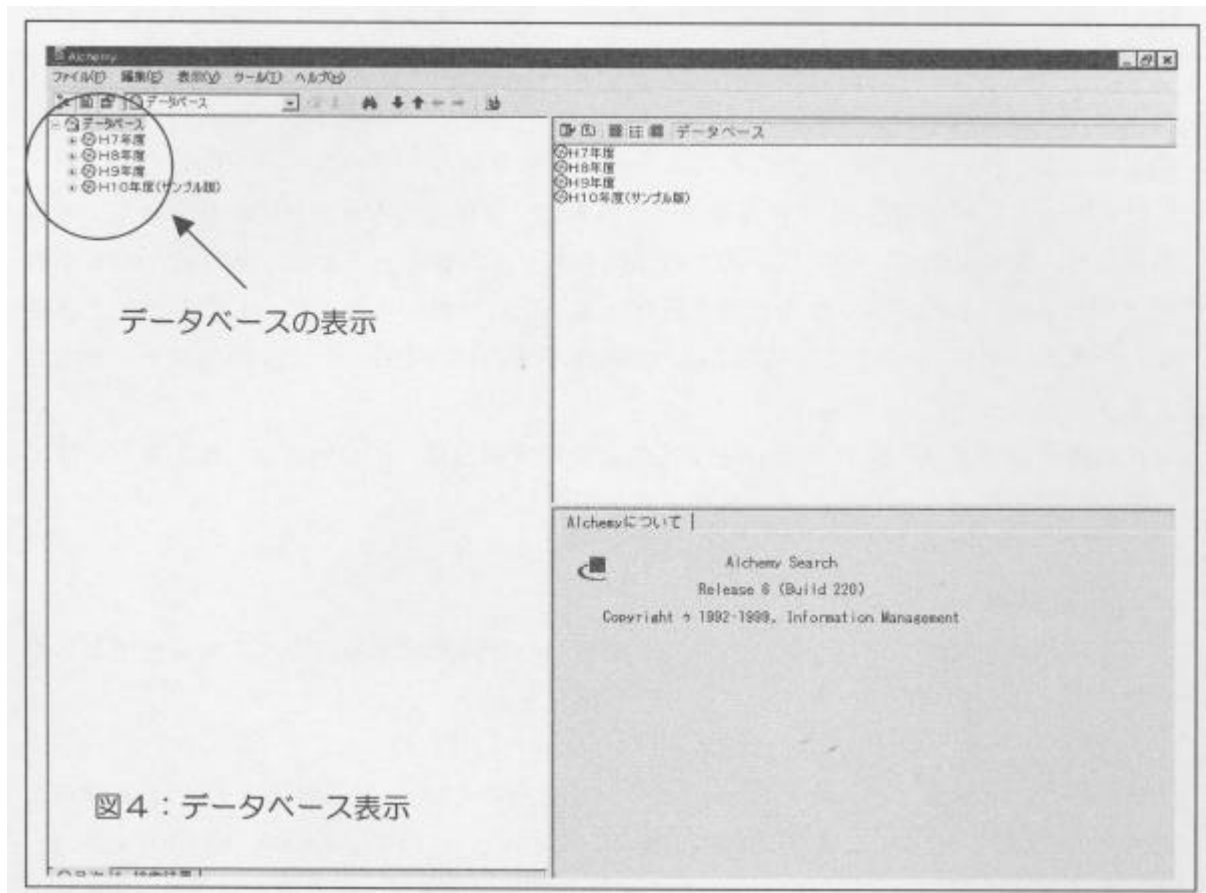


図4：データベース表示

データベースを読み込むと、図4の画面になる。この画面では、平成7年度、平成8年度、平成9年度、平成10年度(サンプル版)データベースが読み込まれていることを表している。

3)各年度の主任研究班と階層構造(図5、6)

各データベースの目次は、各年度の主任研究班の研究テーマ、分担研究者の研究テーマ、研究協力者の研究テーマごとに、階層構造になっており、主任研究班の下の階層に分担研究班があり、その下の階層に研究協力者が表示される。

4)次に、検索機能を検証してみる。

各データベースの目次(主任研究班)が表示された状態で、メニューバーの検索メニューまたは検索アイコンを開いてみると、図6のようになる。ドキュメント情報とドキュメント内容の両者を

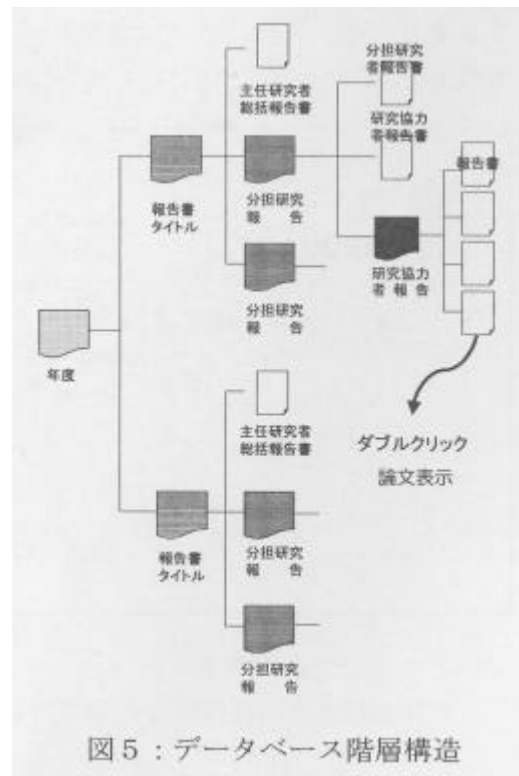
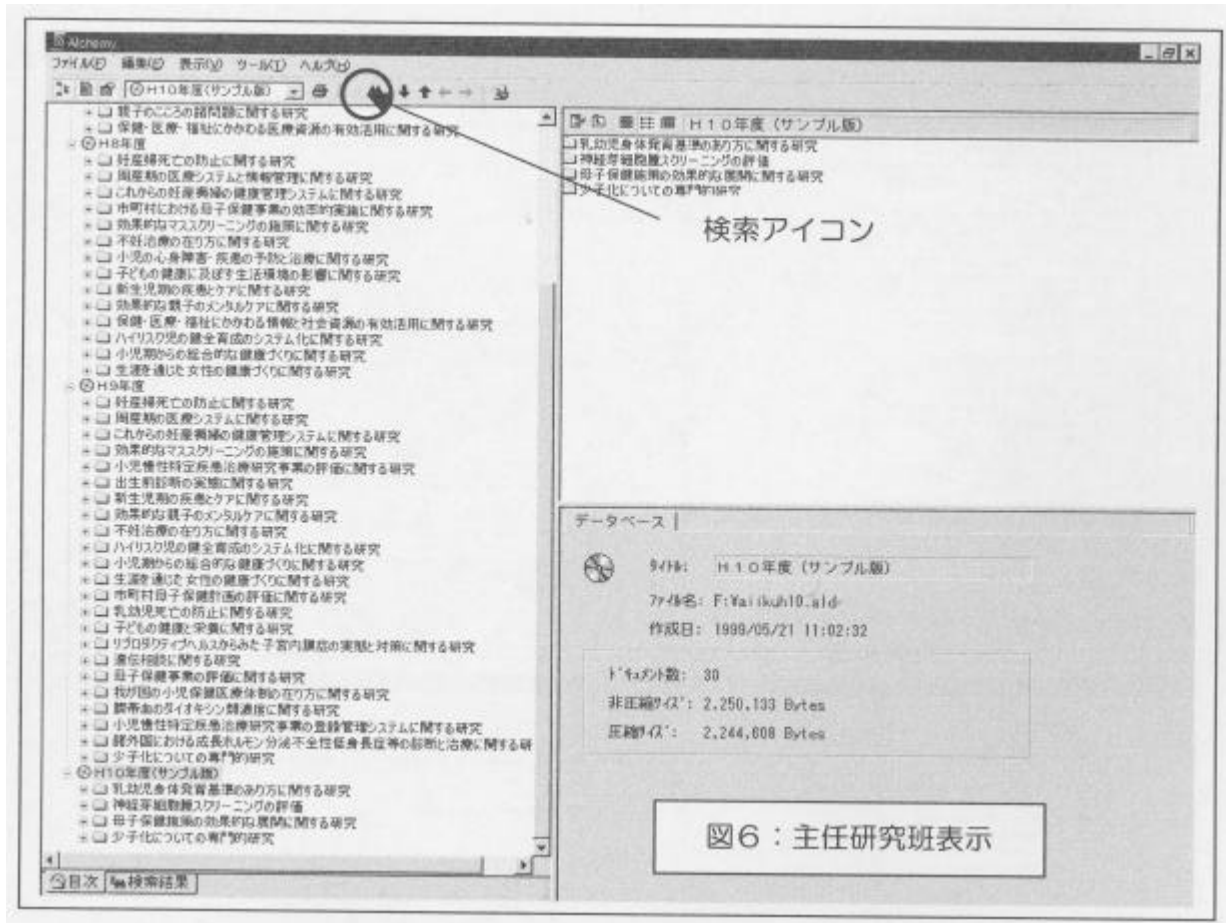


図5：データベース階層構造

検索の対象とするように、チェックボタンにチェックを入れておく。ドキュメント情報フィールドとは、各研究タイトル、著者、キーワードを意味しており、る。ドキュメント内容とは、要約、目的、まとめなどOCRで切り出しテキスト化して、論文の最終ページにリンクしてあるテキスト文を意味する。また、新規分では、論文の7~8割は全文テキストで取り込み、PDFファイル化してあるので、全文の文字列検索が可能である。



次に、図7に示したように、検索したい文字列を検索式のフィールドに入力する。検索にはアスタリスク(*)やワルドカード(?)が使用できる。即ち、AAA*は先頭一致を意味する。全ての文中からAAAという文字列を検索したいのなら、*AAA*を用いる。また、ANDとORの演算子を用いることもできる。

例題として、*子育て*and*支援*で検索してみる。文中に「子育て」と「支援」という2つの単語がある報告書を検索することを意味している。

結果は、図8に示した。各データベースごとに抽出されたドキュメントの数と研究報告書のテーマを表示している。

5) 報告書内容の表示と目次との関連づけ

検索されたドキュメントの先頭のがハイライトされて、右側の上段のウィンドウに報告書の内容がPDFファイルで表示される。

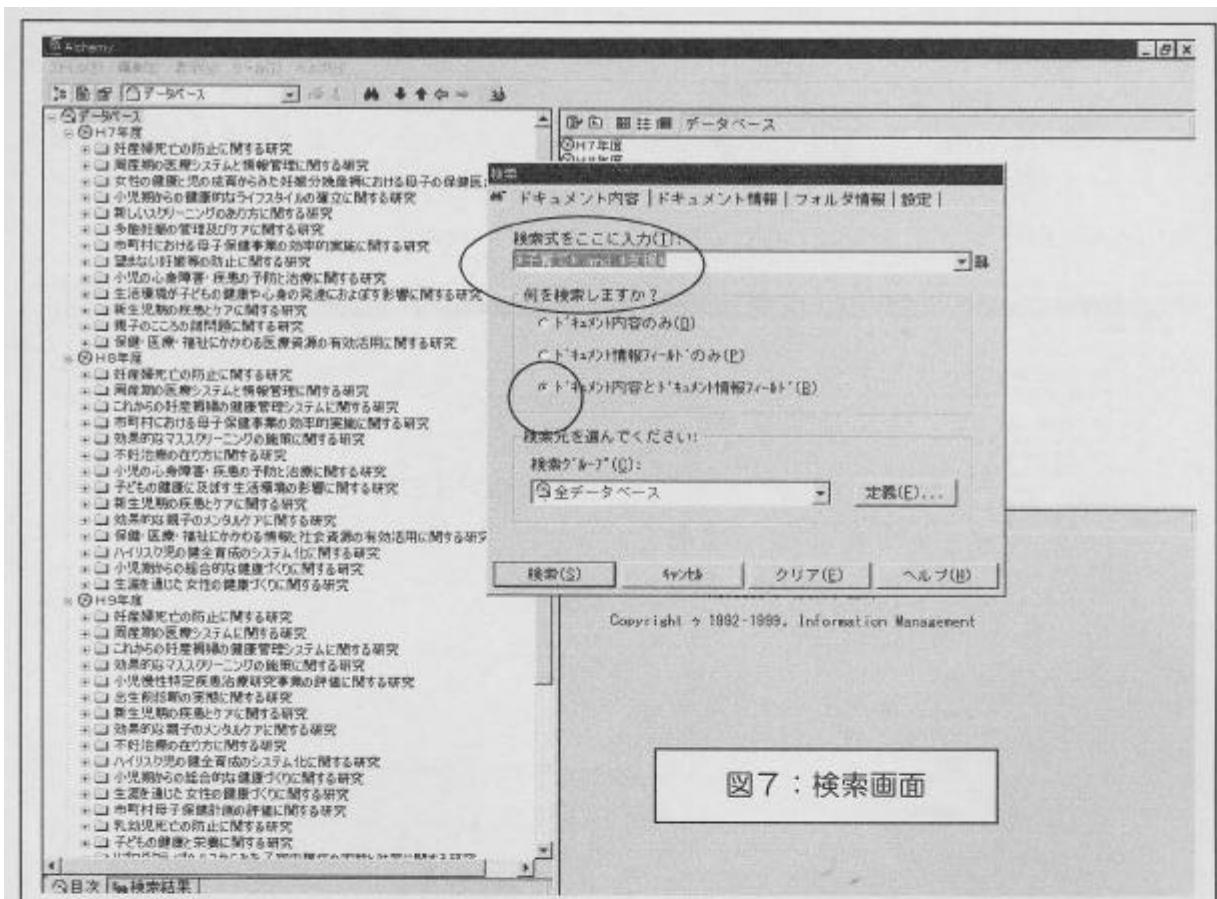
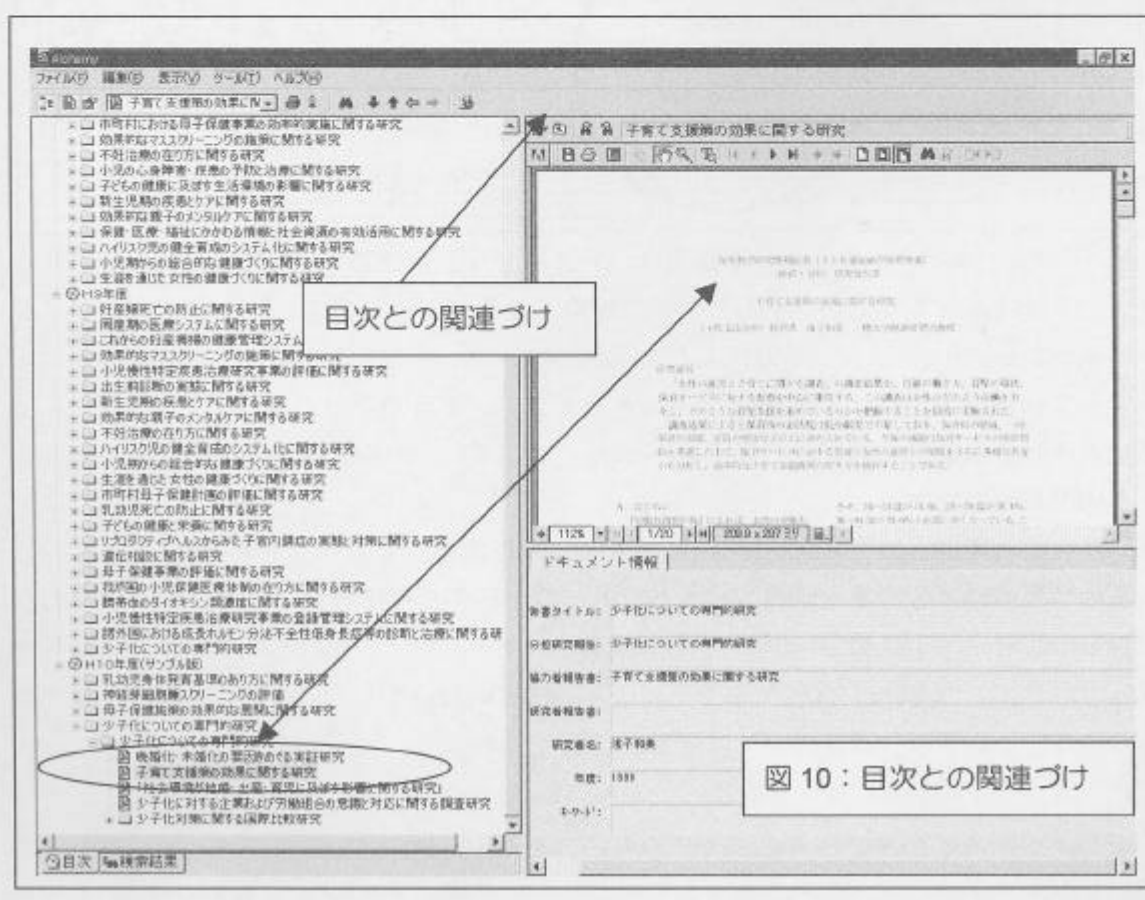
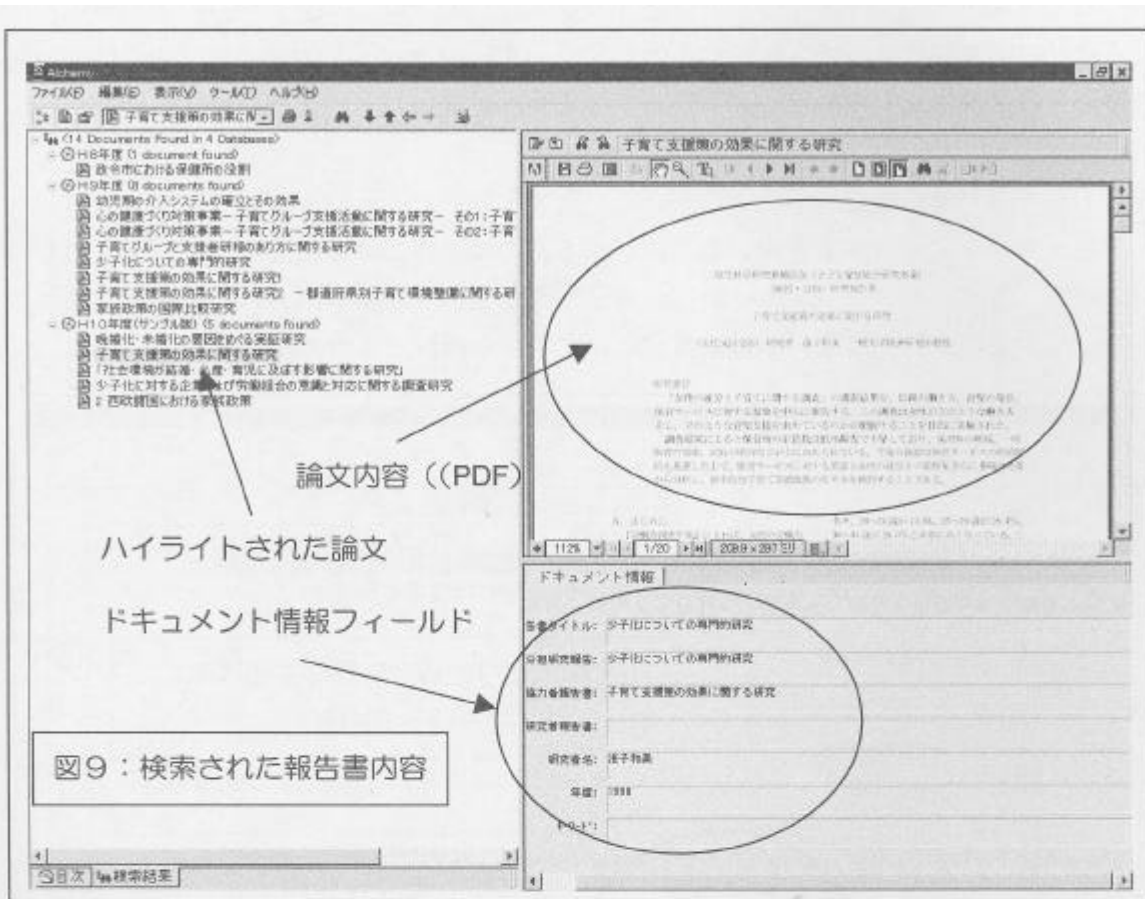


図7：検索画面



図8：検索されたドキュメント

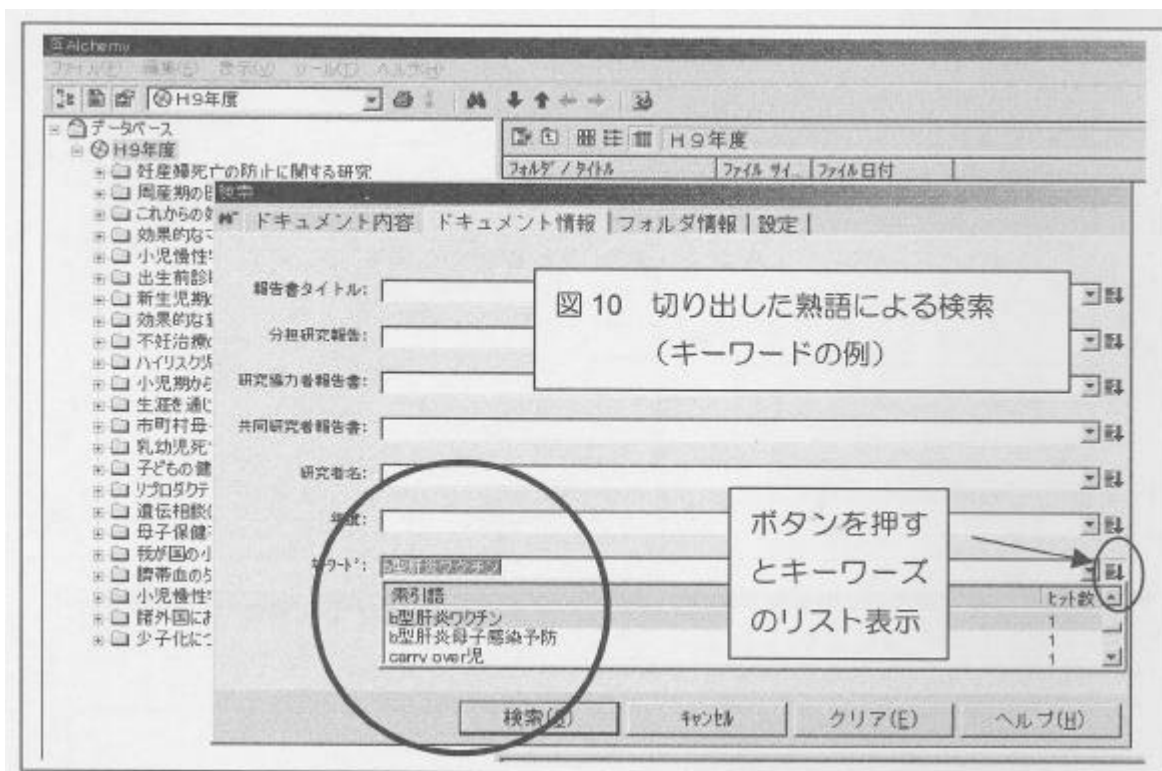


目次との関連づけアイコンをクリックすると、PDF ファイルで表示されている論文が報告書のどの階

層に属しているかがわかる。つまり、研究班のどの位置に属しているものかがわかる。

6) もう一つの検索機能

ドキュメントフィールドを用いる方法がある。この方法では、任意の文字列を入力してもよいが、各ドキュメントフィールドの表示ボタンを押すと切り出された文字列が表示される。この表示された文字列から検索したい文字列を選択して検索ができる。

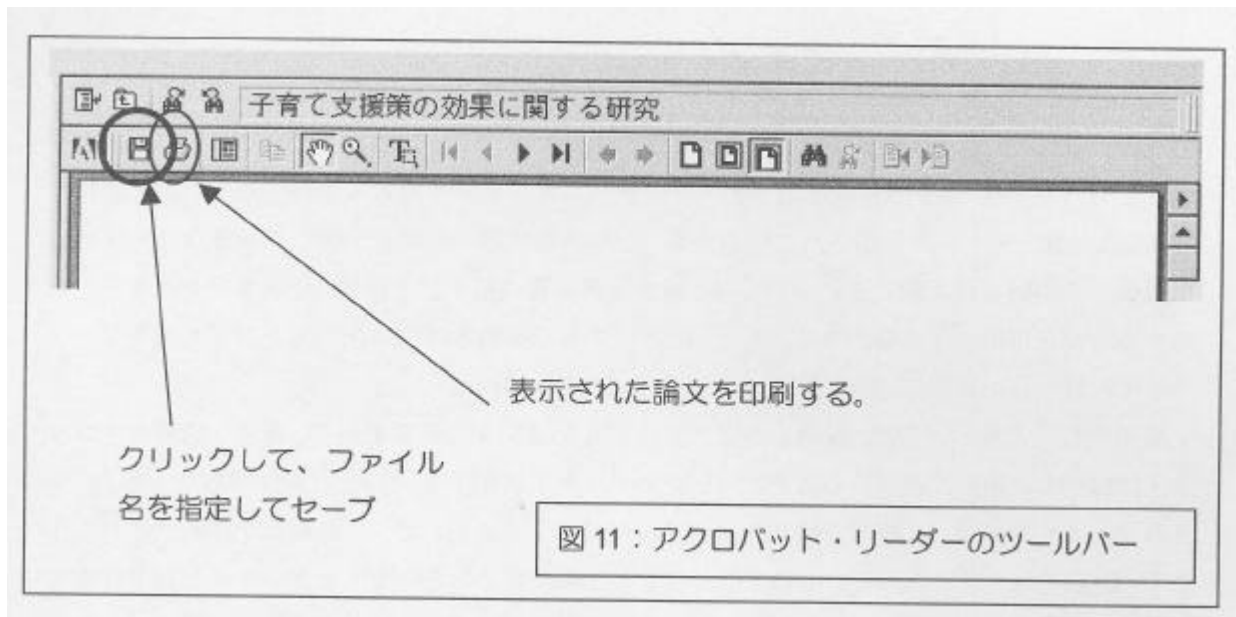


PDF ファイルで表示されている論文は、アクロバットの「ファイルアイコン」をクリックし、ファイルとしてセーブすることができる。もう一つの方法は、Alchemy のファイルメニューから、「ファイルの取り出し」を選択して、ファイルをセーブするドライブとディレクトリーを指定すれば、PDF ファイルとしてセーブすることができる。

7) 表示された論文の印刷は、アクロバット・リーダーの印刷アイコンをクリックすることで印刷が可能である。

8) 最後に、印刷された PDF ファイルのサンプルを添付しておく(巻末資料)。

このサンプルは報告書を画像ファイルとして取り込んだ過去分データのサンプルである。報告書の最終ページに OCR で取り込み、検索用の文字列として添付したテキスト文章が付属している。この部分は、OCR ソフトにより、機械的に処理しているので、誤字や脱字のみられることがあり、取り扱いには注意が必要である。



D. 考察

過去の「心身障害研究」および今年度から実施される厚生科学研究「子ども家庭総合研究」の成果が広く有効に活用されるためには、より多くの関係者に報告書が配布され、研究内容を手軽に入手できることが重要である。

従来 of 厚生省心身障害研究報告書は優れた研究成果が掲載されているにも関わらず、限られた行政機関や研究機関に配布されたのみで、多くの学徒、研究者、実地で活躍する医療・保健・福祉の専門家の目に触れにくいという難点を抱えていた。

そこで、これらの報告書の普及と有効活用を図るためには、報告書を電子化し、検索機能を備えたデータベースを構築するのが最も有効な方法であると考えられた。電子化された報告書は、従来の印刷物に比べ、コンパクトな媒体に大量のデータを収納することが可能であり、いったん、電子化すれば需要に合わせて、その後の複製も容易であり、保管場所などの制約と増刷に多くの費用を要する印刷物に比べれば、対費用効果も優れている。また、データベース化により、大量の論文の中から、必要な資料を検索することができるなど、印刷物で保管するより遥かに利便性が高いと考えられる。

今年度は、昨年度検討し確立した電子化の方法論にしたがって、実際のデータベース化作業に着手した。また、若干、技術的問題を残してはいるが、十分実用に供することができるデータベースに仕上がっている。

昨年度研究の要約を加えながら今年度研究の成果と問題点について述べると次のようになる。

1) 電子データ化の方法論

報告書の電子データ化にあたって、既に製本されている過去分「心身障害研究報告書」とこれからの「子ども家庭総合研究報告書」に分けて検討した。

過去分の心身障害研究報告書については、電子化作業の効率と経済性から、画像データを使用したデータベース作成が妥当であると考えられた。方法論は報告書を解体し論文の各ページをイメージキャ

ナーで読みとり、画像ファイルとして保管する方法を用いた。

新規の「子ども家庭総合研究報告書」は、7 から 8 割は電子データとして収集が可能であると考えられる。しかし、そのファイル形式は多様な機種やソフトで作成されていることは過去の研究でも明白であり 1)、その多様な電子データを統一したファイル形式に変換することが必要になる。また、電子データで提出されない報告書、即ち印刷物のみの場合や変換不可能なワープロ専用機などで作成された報告書については、過去分報告書の電子化と同様に、画像データとして収録する方法を併用する必要がある。また、指定ソフトで作成されていない図表なども画像データとして収録し、当該論文ファイルとリンクさせる必要がある。

電子化した大量データは、収納するファイル形式を選択する必要がある。多くの情報をコンパクトに収納し、利便性を高めるためには、ファイルサイズが小さいことも条件であり、電子データのファイル形式は、現時点では PDF ファイルが最適と考えられる。これはファイルサイズが小さく、印刷イメージと近似した形態で収録でき、どんなプリンターでもオリジナルどおり印刷でき、テキスト部分・イメージ部分ともに電子的切り貼りが可能である。また、データベース化を図るとき、PDF ファイルに収録したテキスト部分は、文字列として検索が可能であり、PDF ファイルに対応した検索ソフトを使用すれば、ほぼ全文検索の機能を備えることができる。

また、PDF を表示・印刷するビューワソフト(Acrobat Reader)は、各種のプラットフォームで閲覧することが可能であり、イメージファイルも PDF に変換することによりファイルサイズが小さくなり、大量データの収録が可能になる。

2) 検索用ソフトの選定

データベースを構築する上で、ポイントになるのはオーサリングと検索機能であり、昨年度の検討結果を踏まえ、Alchemy release 6 を用いて実用化に着手した。

Alchemy は、現在さらに機能アップし、ODBC ドライバーに対応し、Visual Basic などを用いたカスタマイズ機能が充実した Version 6 がリリースされ、日本語バージョンも完成された。しかしながら、問題もあり、わが国で医療関係機関を中心に普及している Macintosh OS には対応しておらず、今後の検討課題として残っている。

3) 今年度の研究成果と問題点

今回、Alchemy Release 6 日本語バージョンにより、PDF ファイル化した報告書を用いてデータベースを作成した。過去分心身障害研究報告書は昭和 50 年～平成 9 年度までの 23 年間についてデータベース化を行い、平成 10 年度新規分子子ども家庭総合研究報告書のデータベース化に着手しており、ほぼ完成の域に達している。

研究報告書のデータベース化に当たって、もっとも重要な機能は、自由語での高い検索機能である。さらに、報告書という印刷物の電子化であり、電子化された書物という体裁も必要になる。

今回作成した Alchemy によるデータベースは、報告書の各論文を、主任研究班、分担研究班、研究協力者、さらに共同研究者などの階層構造に分けて、比較的自由に設定されており、報告書の目次のイメージを再現できている点で、利用者にとって原本の報告書と変わらぬ自然な構造で構築できている。検

検索機能には、多少問題はあるが、検索文字列を演算子で組み合わせることで検索することにより、必要な論文を探し当てることができる。また、数年度のデータベースを合体して同じ文字列の検索を行っても、検索のスピードの低下は気にならなかった。

過去分報告書は、報告書そのものは画像データで電子化しているため、検索の対象になる文字列は、論文のタイトル、著者、キーワード、OCR で切り出したテキスト部分であり、十分な検索機能を補完しているとは言い難い。しかしながら、既製本されている印刷物を用いた電子化であり、現時点ではここまですべてが技術的な限界である。

新規分報告書では、もともと電子データで提出されたものを用いて、PDF ファイル化を行っており、7～8割の報告書はこの方法で対応可能である。しかしながら、問題もあり、報告書の作成に用いられたソフトが多様であり、過去分と同様にイメージで取り込まなければならない報告書が全体の約2～3割を占めている。また、医療機関を中心としたマッキントッシュユーザーも多いため、マッキントッシュ版ソフトを多数用意せざるを得なかった。

Windows ソフト	Macintosh ソフト
Microsoft Word	Microsoft Word
Microsoft Excel	Microsoft Excel
一太郎	一太郎
その他、テキストファイル、BMP ファイルなど	Canvas
	EGWORD
	Word Perfect
	Page Maker
	その他テキストファイル

新規分報告書データベースでは、ほぼ全文検索が可能であり、検索機能は強力である。しかしながら、将来的に問題がないわけではなく、全文のすべての文字列を検索対象とするため、文字列を探し出す作業に時間を要し、体感的に検索スピードの遅さが問題になる可能性がある。今回検討したサンプル版では、データ数が少ないためか、体感的検索スピードに問題は感じられなかった。

別添資料：厚生省心身障害研究報告書 CD 化リスト一部報告書の欠損がある。

【文献】

- 1) 齊藤進、庄司順一、中村敬、恒次欽也、中沢明紀：心身障害研究報告書のデータベース化に関する研究、厚生省心身障害研究「保健・医療・福祉にかかわる情報と社会資源の有効活用に関する研究平成8年度研究報告書,p158～163,1997
- 2) 広田健一郎：日本語 PDF + Acrobat 入門,P15～46,1997.7.5
- 3) 中村敬、庄司純一、恒次欽也、中沢明紀：心身障害研究および子ども家庭総合研究事業報告書のデータベース化に関する研究、平成10年度厚生科学研究(子ども家庭総合研究事業)報告書(第6/6):189～195,平成11年3月

市町村母子保健計画の推進方針に関する研究
心の健康づくり対策事業—子育てグループ支援活動に関する研究—
その1：子育てグループの実態について

中村 敏¹ 小山 修² 斎藤 達³ 樋口美奈³ 長坂典子³ 高野 隆³

【要約】近年、従来の地域的住民活動とは別に都市型の地域活動として、子育てグループの活動が盛んになってきている。この活動は、一部行政の支援の影響もあって、全国的に広まりつつある。活動の特徴は同じ目的意識をもちたもの同士の自発的活動とみることができ、今回は、いくつかの調査をまとめて、その実態の姿を掘り下げてみた。このグループ活動は、地域性が強く、比較的年齢の近いもの、子育てという同じ境遇におかれたもの同士が集まり、趣味的活動であったり、公共的活動であったり多種多様な活動を行っている。

これらの多くのグループは、単にメンバーが集まることで、目的が達成されていることも多く、仲良しグループ的な要素も大きい。また、中には組織化され支部をもつ大きなグループに発展しているものもあり、規模や成り立ちも様々である。このグループ活動は、その背景によりいくつかに分類することができる。最も多いのは、一般の母と子どものグループであり、比較的子ども同士の年齢が近いもの同士が集まって結成されており、自主的な活動を展開しているものが多い。

【見出し語】
子育てグループ 実態 活動内容

【目的】本格的な少子化時代を迎え、社会を 社会運常意識の希薄化により、地域の要請を上げての育児支援が求められている。育児と 能が低下してきており、社会全体での子育て支援の推進がある保健・医療、福祉、教育の 支援が求められている。

各部門は、縦割り行政の中で発展を遂げ、大 厚生省は、市町村が実施する「子どもの心 かな成果を上げてきたが、近年になり各部門 の健康づくり対策」事業として、心の健康づく 同士の結ぶ関係の弱さが大きな社会問題に な くり運動の推進、育児不安への対策、虐待・ っている。また、近年、少子化、核家族化、 いじめ対策を挙げ、この中の育児不安への対

¹ 母子養育会白木子ども発達総合研究所
² 東京都立大学大学院研究科
³ 母子養育会総合母子保健センター研機部

検索用テキスト OCR(光学的文字認識)ソフト使用
論文の一部ですが、図解等の関係で断片が含まれる場合があります

【要約】近年、従来の地域的住民活動とは別に都市型の地域活動として、子育てグループの活動が盛んになってきている。この活動は、一部行政の支援の影響もあって、全国的に広まりつつある。活動の特徴は同じ目的意識をもちたもの同士の自発的活動とみることができ、今回は、いくつかの調査をまとめて、その実態の姿を掘り下げてみた。このグループ活動は、地域性が強く、比較的年齢の近いもの、子育てという同じ境遇におかれたもの同士が集まり、趣味的活動であったり、公共的活動であったり多種多様な活動を行っている。

これらの多くのグループは、単にメンバーが集まることで、目的が達成されていることも多く、仲良しグループ的な要素も大きい。また、中には組織化され支部をもつ大きなグループに発展しているものもあり、規模や成り立ちも様々である。このグループ活動は、その背景によりいくつかに分類することができる。最も多いのは、一般の母と子どものグループであり、比較的子ども同士の年齢が近いもの同士が集まって結成されており、自主的な活動を展開しているものが多い。

この画面は検索した論文を印刷したものの
要約の部分をOCRで取り込みテキスト化した
したもの、文字列検索の対象になる。