

図書分類エキスパート・システム

細野公男、田村俊作、原田隆史（慶應義塾大学文学部）
諸橋正幸、梅田茂樹、隅田英一郎（日本IBM東京基礎研究所）

図書館では、複雑であいまいさを含んだ作業が比較的多く、その円滑な遂行には問題領域に関する各人の専門知識および長年培った経験的知識を必要とする場合がしばしば生じる。したがって、例えば、図書の分類作業の効率化、高度化をはかるには、図書館員の広範な専門知識と実務体験を通して得た経験則を十分に組み込んだエキスパート・システムの構築がきわめて効果的である。この点に鑑みて、著者等は現在図書分類エキスパート・システムの研究・開発を行っている。

このエキスパート・システムでは、図書のタイトル、目次、はしがきの3つの情報源から人間が抽出したキーワードに基づいて分類番号が決定される。使用する分類表は、階層構造を持つ日本十進分類表（NDC）であるが、NDCの階層にはゆがみがあるので、誤った分類がなされないために廣NDCが作成されている。したがって、NDC分類番号の参照はこれを経由して行われる。つまり、キーワードを廣NDC番号に変換するルールと廣NDC番号をNDC番号に変換するルールとによって分類番号が生成されるのである。また、情報源の違いによってキーワードの影響力が異なるようにするため、確信度の値をキーワードの出現場所によって変化させている（タイトルは0.8、目次は0.6、はしがきは0.5）。さらに、キーワードが複数の情報源にまたがって出現する場合は、対応する分類番号の重みを高くし、階層上同一レベルのキーワードが多数出現する場合は、それぞれの分類番号よりも上位の番号の重みが高くなるようにしている。

このような特徴を持つ本システムによって、分類担当者の手ですでに分類済みの保険分野の図書41冊を対象に分類を行ったところ、分類担当者による結果と一致した件数は31件であり、一応の成果が得られた。

Developing an Expert System for Book Classification

Kimio Hosono, Shunsaku Tamura, Takashi Harada
School of Library and Information Science, Keio University
2-15-45 Mita Minato-ku, Tokyo
Masayuki Morohashi, Shigeki Umeda, Eiichiro Sumita
Tokyo Research Laboratory, IBM Japan Ltd.
5-19 Sanbancho Chiyoda-ku, Tokyo

This paper describes the on-going project to develop an expert system for book classification. After having acquired classifiers' expertise by using a protocol analysis method, production rules for selecting class number(s) of the books from the Nippon Decimal Classification Schedule are composed.

The process to determine class numbers is as follows.

- a) keywords are extracted from the title, contents and preface of books and their degree of importance is calculated based on the place of their occurrence.
- b) those keywords are integrated and translated to class numbers by using the relative index and main table of the Classification Schedule.
- c) class numbers are outputted by decreasing order of the certainty.

At an experiment 41 books in the field of property(damage) insurance are classified and 31 of them are successfully classified.

1. はじめに

情報化社会の進展を反映して、図書などの形態をとる記録情報は増加の一途を辿っており、また、その一方で情報ニーズの高度化、多様化が顕著になっている。その結果、代表的な情報提供機関の1つである図書館に対する要求は、益々高度になりつつある。図書館業務の機械化、システム化が広く行われているのは、このような社会的要請に的確に対処し、応えるためである。

図書館での作業の多くは、ある定められた規則、手順に則って遂行されているが、その手順、規則は必ずしも完全に規定されたものではない。また、複雑であいまいさを含んでいるため、論理的なアルゴリズムを見出すことが困難な作業も比較的多い。その典型的な例が図書などの目録・分類作業、情報要求者に対するレファレンス・サービスの中に見られる。そのような場合、図書館員は問題領域に関する各人の専門知識および長年培った経験的知識に基づいて、意思決定を行わねばならない。これは、意思決定の結果に図書館員間で差異が生じること、また、担当する図書館員が豊富な知識と経験を持てば持つほど、その結果の信頼性が高まることを示している。

したがって、単なる機械的なデータ処理では、この種の業務の効率化、高度化をはかることは困難であり、図書館員の広範な専門知識と実務体験を通して得た経験則を十分に組み込んだ機械化システムの構築が必要である。この点で、エキスパート・システムの導入は、図書館業務・サービスの高度化に資するところが大きい。そのため、近年、欧米をはじめとして図書館業務へのエキスパート・システム導入に関していくつかの試みがなされており、この傾向は年々強くなっている。

そこで、本稿では図書館業務を対象とした代表的なエキスパート・システムの例を紹介すると共に、著者等が現在研究・開発中の、『日本十進分類表』に基づく図書分類エキスパート・システムを報告する。

2. 図書館業務を対象としたエキスパート・システム例

a. 導入への試み

エキスパート・システムを導入する試みは、1984年頃から開始された。そして現在までは、目録作業のようにルールに基づいて多くの複雑な選択肢の中から最適なものを選ぶことが要求される作業と、情報提供サービスにおけるレファレンス・インタビューのようにユーザーのあいまいで不完全な情報ニーズを明確かつ具体的にする作業とに、その対象を大きく分けることができる。なお、情報提供サービスでは、検索の方法・手順の決定に関しても本稿で言及した種類の知識を必要とするが、エキスパート・システムの導入がはかられているのは、ユーザーの情報要求の把握を対象とするものが多い。前者に属する例として、Hjerpe¹⁾、Davies & James²⁾、Davies³⁾、Burton⁴⁾、Borko⁵⁾などが、後者の例として、Vickery & Brooks⁶⁾、Pollitt⁷⁾などがあげられる。なお、我が国でも開発例がある⁸⁾。

b. PLEXUSシステム⁶⁾

エキスパート・システムの導入に関しては、aで示した以外にもいろいろの試みがなされており、全体としてその数は増加しつつある。しかし、aの例を含めまだ多くは考え方、基礎設計の段階にある。その中でプロトタイプ作成が終了し、一応稼働の段階に達しているのがこのシステムである。

これは、ロンドン大学の Central Information Service (CIS) が公共図書館での情報提供サービスを対象にして開発中のシステムで、ユーザーから提示される圏芸に関する質問に答え得る情報源、つまり場所、専門家、図書、機関・団体の情報を提供する。使用言語はPASCALと PROLOGで、SIRIUS I マイクロコンピュータ (内部記憶容量 842K バイト、20メガバイトのハード・ディスク) 上で実現されている。このシステムは以下のモジュールから構成される。

i) ユーザー・モデルの作成

情報質問を受け付ける前に、PLEXUSシステムの精通度、園芸の経験度、この種の問題を解決するために過去どのような処置をとったことがあるか、ユーザーの居住地域、園芸上の問題が発生した地域などの項目について、ユーザーに質問する。これはPLEXUSが提供するメッセージのレベル、情報の種類・レベルを決定するのに使われる。

ii) 質問の処理

質問は255文字を越えない範囲の自然語文で構成される。入力された質問から不要語が除かれ、残った重要語の語尾処理が行われ語幹にまとめられる。これらの語幹はシステム辞書と照合され、それが属する意味カテゴリー、BSO分類表中のクラス、同義語グループの番号と共に一時ファイルに蓄積される。ユーザーの質問内容は、必ずしも回答を与えるのに十分な情報を含んでいるわけではない。そこで、情報の不足の有無を調べ補足するために各意味カテゴリー毎にフレームが用意されており、スロット（その意味カテゴリーと概念的に関連する他の意味カテゴリー群で構成される）に値がセットされているか否かで、質問の妥当性をチェックする。したがって、質問を構成する各重要語の属する意味カテゴリー・フレームのスロットが他のフレームからの情報でうめられない場合、ユーザーから追加の情報を求めることになる。同形異義語の処理は質問中での語間の近接度に着目するプロダクション・ルールによって行われる。エキスパート・システムの観点からみた場合のPLEXUSの特徴は、質問の処理モジュール（つまり、ユーザーの情報要求の把握）に専門家の知識が導入されていることである。

iii) 検索

検索は、質問処理モジュールで最終的に得られた重要語間のブール演算で行われる。つまり、同義語間はOR、それ以外はAND関係で処理される。検索件数がきわめて少ない場合は、BSO分類表上で上位にある語を代りに使う、重要語の一

部を検索式から削除する、などの工夫がなされている。検索結果の表示など検索以後の情報処理は、既存の検索システムと本質的な差異はない。

3. 図書分類のためのエキスパート・システム

現在著者等は、階層構造を持つ『日本十進分類表 (Nippon Decimal Classification)』（以下NDCと略称する）に基づいて“保険 (NDC番号は339)”分野の図書を分類するエキスパート・システムを開発中である。このシステムは、IBMで開発されたエキスパート・システム構築ツールであるExpert System Environment (ESE)を使用しており、知識はプロダクション・ルールで表現されている。推論方式は、後向き推論を使用している。なお、ESEでは日本語が扱えないので、キーワードの表記はローマ字である。

a. エキスパート・システム構築の意義

分類作業は分類担当者によっておこなわれる。図書の分類にあたっては、図書の主題を理解しその構成概念を抽出するのに必要な知識（多分に経験的）、分類表の構造や使い方に関する基本的な知識、そして、主題構成概念に分類表中の分類記号を対応させる経験的知識・技術が不可欠である。このような専門知識、実務経験の差異によって、分類担当者間で分類番号の付与に一貫性が欠け同一の図書に異なった分類記号が無秩序に付与されるような事態が生じるのは、適切なアクセス・キーの提供という観点から考えると好ましくない。また、分類されるべき図書の量は増大の一途を辿っており、分類作業に要する時間は著しく増大している。このような環境下で高品質の分類結果を比較的短時間で得るためには、高いレベルの知識・技術を持つ分類担当者を多く必要とするが、図書館が手当てできる人数には限りがある。したがって、エキスパート・システムの導入は、迅速かつ効率的でしかも均質な分類結果を得るための1手段として、きわめて有望であるといえよう。

b. 図書の分類作業

図書の分類とは、個々の図書にその図書の主題を表わす分類記号を付与する作業であり、その目的は、図書の検索、利用を容易にするためのアクセス・キーを提供することである。図書を検索するためのアクセス・キーとして、著者名や書名などの他にその図書の主題（内容）を表現する“ことば”や分類記号があるが、分類作業は分類記号の形態をとるアクセス・キーを提供する。

分類作業は、以下のプロセスに分解されると考えられる。しかし、人間が行う場合は、このプロセスが明確に区別されているわけではなく、渾然一体となっている。

- 1) 分類対象である図書の主題を理解、把握する。
- 2) 主題を構成すると考えられる重要概念を抽出する。
- 3) 分類表の使用指針、分類規準に基づいて上述の主題構成概念に対応する分類記号（番号）を分類表から選択し、図書に付与する。このプロセスでは相関索引を介して、抽出された重要概念に対応する分類番号を選択することが多いといわれている。

分類規準の一般的な通則として、わが国の代表的な分類表であるNDCでは、例えば次のようなものをあげている⁹⁾。なお、これに加えて、各図書館毎に固有の規準が設定されている。

i) 主題と形式

まず、主題によって分類し、次いで必要があればその主題を表現する形式（理論、歴史的・地域的論述、参考意図書、論文・講演集、逐次刊行物、学会、研究・指導法、叢書）によって細分する

ii) 多くの主題を並列的に扱った図書は、それらを含む上位のクラスの下に分類する

NDCは、分類本表、補助表、相関索引から構成されている。本表は階層構造をとっており、階層上同レベルの概念は同一桁数の数字でその分類番号が表わされるのが原則である。相関索引は、

重要概念を表現するキーワードから分類番号を探す索引である。また、補助表は主題をいろいろの側面（例えば、表現形式、地域、時代等）でさらに細分するための表である。

4. 分類担当者の分類作業の把握

上述した分類作業のプロセスを具体的にルールで表現するためには、分類担当者が図書を分類する際にどのような行動をとるかを理解しなければならない。しかし、分類作業の方法・手順、その結果の評価などに関する研究は、これまでもいろいろなされているが、分類作業中に分類担当者がどのような情報処理を行っているかについての研究はほとんどなされていないのが実情である。例えば、図書内容の把握に関しても表面的な作業手順は明らかにされているが、その過程で得られた情報が分類記号の決定にどのように処理・利用されるかは解明されていないのである。つまり、分類作業は経験的なノウハウに依存するのとらえ方が強く、その結果、分類作業マニュアルが整備されている図書館はほとんどないといわれている。

そこで、まずそれを明らかにするために人文・社会科学分野の図書の分類に携わる図書館員の分類行動をプロトコル分析等によって調査した。図1はその結果である¹⁰⁾。

文学以外の図書を分類する場合は、まずその図書の主題を表わすと思われる語を図書中から抽出する。その際の情報源を重要度の順にあげると、タイトル・グループ（タイトル、サブタイトル、シリーズ名）、帯グループ（帯、カバー）、目次グループ（まえ書、目次、あと書）、著者グループ（著者名、出版者名）、本文グループ（本文、索引、参考・引用文献）であった。つまり、分類担当者は、このような情報源からキーワードを抽出し、相関索引をひいて対応する分類番号の決定を行うのであり、その際、キーワードの重要度にグループ間で差異があることを考慮して総合的な

判断を行うと考えられる。

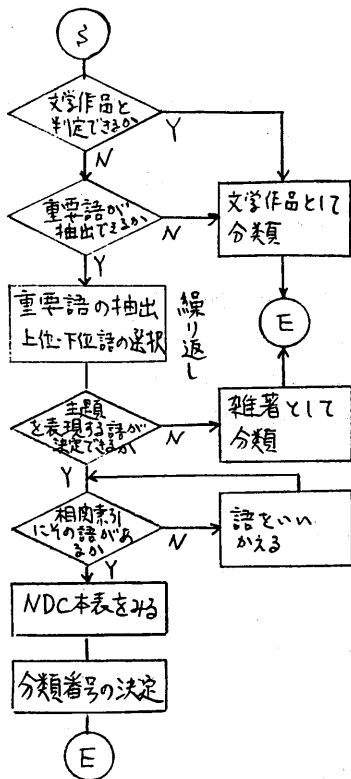


図1 分類番号決定までの流れ

5. エキスパート・システムの構築

a. 分類番号の決定プロセス

本システムでは、調査結果を踏まえて図2のような分類番号決定プロセスを考えた。その具体的な特徴は、以下のように集約できる。

1) タイトル・グループ、目次、はしがきを情報源として、それぞれの情報源から人間がキーワードを抽出しシステムに入力する。はしがきでは、図書の主題を記述していると判断できる、例えば、「本書は・・・」のような形態の文に着目してキーワードを抽出する。

2) 入力されたキーワードに対する分類番号は、主題を表わす数字と表現形式、地理区分などの側

面を示す数字とをつなぎ合せたかたちで表現する。

3) 分類番号は後述する2種類のルールによって生成する。

4) 情報源の違いによってキーワードの影響力が異なるようにするため、確信度の値をキーワードの出現場所によって変化させる。

5) 分類番号は確信度の降順に出力する。

6) キーワードが複数の情報源にまたがって出現する場合は、対応する分類番号の重みを高くする。

7) 階層上同一レベルのキーワードが多数出現する場合は、それぞれの分類番号よりも上位の番号の重みが高くなるようにする。

8) NDC本表には階層のゆがみがあるので、廣NDCを作成し、本表への変換はそれを經由して行う。

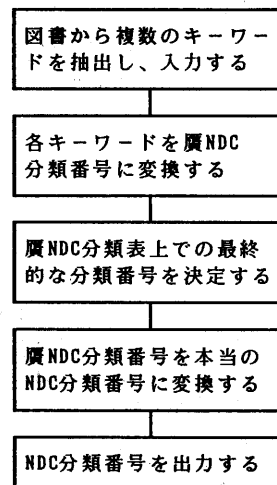


図2 分類番号決定プロセス

b. 廣NDC

NDC本表では、十進分類のため同一レベルのクラスが最大9個しか用意できないこと、また番号の簡潔さを維持するためなどから、番号の長さは必ずしも階層構造を反映していない。その結果階層の表現にゆがみが生じている。例えば、以下

に示すように「339 保険」の中の損害保険の分類番号は、概念階層上では下位に位置するはずの各種保険の番号と同レベルの長さになっている。

- 339.5 損害保険
 - .6 火災保険
 - .7 運送保険
 - .8 海上保険

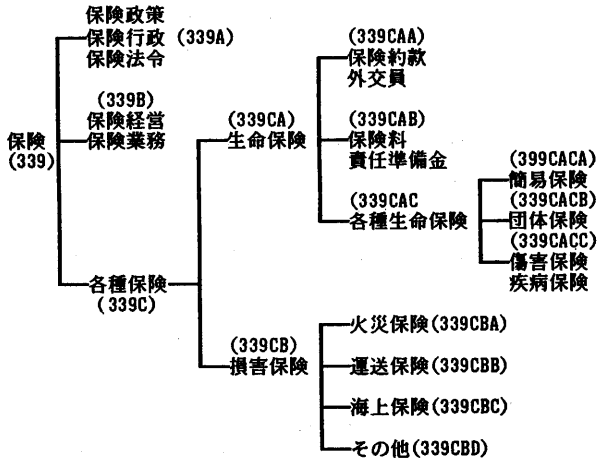


図3 廣NDC

このようなゆがみから誤った分類番号が生成されることを避けるために、廣NDCが導入されている。図3はその一例である。

c. ルール

本システムでは以下のルールが用いられている。

i) 語を分類番号に変換するルール

- ① キーワードを廣NDC番号に変換するルール (相関索引をひくルール)

例 IF TITLE IS 'KASAIHOKEN' THEN
 THERE IS .8 EVIDENCE THAT S1 = '339'
 AND THERE IS .8 EVIDENCE THAT S2 = '339C'
 AND THERE IS .8 EVIDENCE THAT S3 = '339CB'
 AND THERE IS .8 EVIDENCE THAT S3 = '339CBA'

確信度の値はキーワードの出現場所によって異なるようにし、タイトルは 0.8、目次は 0.6、はしがきは 0.5とした。このルールはキーワードに直接対応する分類番号だけでなく、その上位に位置する番号も同時に得られるようにするためである。その結果、入力された複数のキーワードの分類番号に共通する上位の分類番号の重みを高くすることが可能になる。

表1 分類実験結果

書名	分類済みNDC No	システムでのNDC No	判定
新種・自動車保険講座	339.5	339.9/339.5	○
保険概論	339	339	○
保険学概論	339	339	○
損害保険の経済分析	339.5	339.5	○
生保金融の窓から	339.4	339.4	○
利益保険	339.5	339.5	○
アメリカ保険用語辞典	339.033	339.033	○
保険の知識	339	339	○
共栄火災30年史	339.067	339.067	○
保険業界	339.35	339.35	○
欧米の団体保険	339.46	339.46	○
簡易保険入門	339.45	339.45	○
人保険の理論	339	339.4	×
フランスの保険制度史	339.235	339.2	×
近代保険の生成	339.02	339.2	×
喜寿記念加藤由作博士論文集	339.04	339.04	○
保険の現代的課題	339.04	339.04	○
国際保険経営論	339.35	339.35	○
ロイズの保険証券生成史	339.8	339.2	×
イギリスの保険制度史	339.233	339.233	○

② 廣NDC番号をNDC番号に変換するルール

例 IF S3 = '339CB' THEN NDC = '339.5'

- ii) 上述の変換ルールを制御するメタ・ルール
(変換ルールをどのような時にどのような
順序で適用するか等を決定するルール)

d. 分類結果

分類担当者の手ですでに分類済みの保険分野の
図書41冊を対象に、本システムによって分類を行
った結果の一部が表1である。実験全体で分類担
当者による結果と一致した件数は、31件であった。
一致しなかった理由として、以下が考えられる。

① キーワードを分類番号に変換するルールの
不足

② キーワードが必ずしも正しく主題概念を表
現し得ない場合がある

③ 廣NDC番号をNDC番号に変換するル
ールの不完全さ

④ 分類担当者とシステムとの間で分類の深度
(表現形式、地理区分に関して)に差異がある

6. おわりに

本システムをさらに高度化するためには、前述
した問題点の解決だけでなく以下の点でも新たな
取り組みが必要である。

1) 今回の研究では、人間の判断によってキー
ワードの抽出が行われたが、処理の均質性を維持
するには、自動抽出が望ましい。

2) キーワードの重要度の差異は、出現場所 (
タイトル、目次、はしがき) のみに基づいて設定
したが、本来キーワード間には主題の識別力、表
現力に差異があるはずである。この違いを重要度
(あるいはルール) に反映させる工夫が必要である。

3) 分類対象分野の拡大と分類の深度に関する
ローカル・ルールの導入とが必要である。

参考文献

- 1) Hjerppe, R., et al. "Project ESSCAPE -
Expert systems for simple choice of access
points for entries: Applications of arti-
ficial intelligence in cataloging", Linko-
ping, LIBLAB Linkoping University, 1985.
- 2) Davies, R. & James, B. "Towards an expert
system for cataloging: some experiments
based on AACR2", Program, Vol. 18, No.4,
p. 283-97 (1984)
- 3) Davies, R. "Cataloguing as a domain for
an expert system", In: Intelligent infor-
mation systems. New York, Ellis Horwood,
1986. p. 54-77.
- 4) Burton, Paul F. "Expert systems in clas-
sification". In: Expert systems in libra-
ries. London, Taylor Graham, 1986. p.50-66.
- 5) Borko, H. "Getting started in library
expert systems research", Information Pro-
cessing & Management, Vol. 23, No. 2,
p. 81-7(1987)
- 6) Vickery A. and Brooks, H. M. "PLEXUS--The
expert system for referral", Information
Processing & Management, Vol. 23, No. 2,
p. 99-117(1987)
- 7) Pollitt, Steven. "CANSEARCH: An expert
systems approach to document retrieval",
Information Processing & Management, Vol.
23, No. 2, p. 119-38(1987)
- 8) 田中芳彦他. "知的図書館システムILIAS",
Expert Systems, 3(1), p.9-12(1987)
- 9) 日本十進分類法 新訂8版. 東京, 日本図書館
協会, 1978. p.22-23.
- 10) 原田隆史, 原青子. "図書館における分類作
業の分析", Library and Information Science,
No. 24, p. 71-9(1986)