

用途別統合検索におけるガイダンス機能

1 L-4

嶺岸 則宏 田中 聡 高梨 郁子

三菱電機(株) 情報技術総合研究所

1.はじめに

将来の電子図書館システムは、ネットワーク化され、老若男女あらゆる人が利用者として想定される。このため図書情報の検索は、情報検索の専門家でない一般の利用者でも、容易に行なえることが重要である。大規模な総合的図書館や、その他、大学、研究機関等の様々な図書館が電子化・ネットワーク化され、利用者の地域/専門/時間を問わない各種の問合せやシステムへのアクセスが容易になる。

このような総合電子図書館に対して、各電子図書館のある一分野の図書情報を統合した仮想的な専門電子図書館を提案する。関連する複数の電子図書館の連携により、ある専門分野の図書情報を検索する場合には、複数の各図書館サーバを検索せず、この仮想的な専門電子図書館にだけアクセスし、統合された結果を得られる。これにより利用者は複数の図書館の図書資産を目的や用途に合わせて効率良く利用することが可能になる。仮想的な専門電子図書館は、バーチャルインデックスの集まりであるので、通常電子図書館のサービスとして同じサーバ上に構築してもよい。

現在の図書情報検索システムは、システム毎に異なる分類体系で整理されているため、検索者は、各分類体系を十分に理解しなければ目的の図書を効率的に検索できない。ここで提案する用途別統合検索とは、上記の仮想専門電子図書館システムを構築し、個別に提供されている様々な情報を、検索者の理解し易い分類に整理統合することによって、様々な検索目的に柔軟に対応するものである。例えば、医学分野の場合、医学生や医師向けには医学図書館向けの極めて専門的な分類体系によって検索、一般人向けには家庭用医学書の目次を基にした分類体系から検索できるようになる。

しかし、このような、様々な分類体系による検索を提供するためには、あらゆる分類体系に対して対応可能でなければならない。そこで前回、我々は分類体系の階層関係に関連度を定義することで

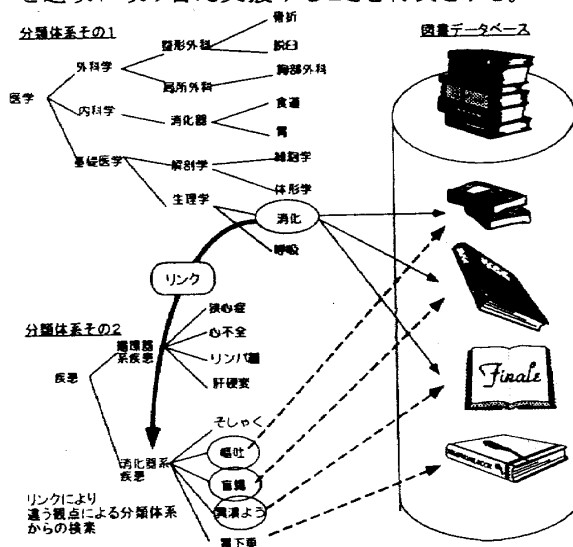
自動的に分類を行う方法について報告した。[1]

前回の評価により、様々な立場の検索者に対応する分類体系による検索機能は実現された。しかし、所詮分類体系を変更するものではなく、例えば「別の観点の分類」「さらに細かい分類」で検索したいと思った場合、別の分類体系の構造を理解して選択し直して最初から検索する必要がある。このような課題に対し、現在検索中の主題に関連する別の観点による分類体系メニューを提示、ガイダンス機能として提供し利便を向上させる事を考えた。

従来、検索ガイダンスの研究は多数なされてる。検索者ごとの検索履歴を利用する手法が代表的であるが、履歴蓄積までに時間がかかる、検索毎に目的が同じとは限らない、等の問題があった。本研究では、それら問題の解決のため用途別統合検索におけるガイダンス機能として、目的や用途毎に作成した観点の異なる分類体系メニューのうち関連する項目を適宜提示し、検索者が選択し切り替えることで実現する。その切替え実現のための様々な分類体系間の関連付けを自動的に行う手法を検討した。

2.ガイダンスデータの作成

本方式のガイダンス機能は、1つの分類体系の中で適切な検索場所を指示したりナビゲーションするのではなく、検索者の検索目的に応じて観点の違う分類体系を適切に切り替え支援することを特長とする。



そのためには、異なる分類体系間で関連のある項目同士を対応づける必要がある。この対応づけに必要なデータをガイダンスデータとし、分類体系とそれらに対応づけられた図書を元に予め自動的に作成する。

対応づけの基本的な考え方を図1に示す。前提として、ある図書群に対して、分類体系1、2それぞれ検索できるような既に分類されインデックスが構築されているものがある。分類体系1で「消化」という項目に分類されている一連の図書があるが、分類体系1ではこれ以上の細分化はされていない。ここで、分類体系1「消化」という項目と関連があると考えられる分類体系2の「消化器系疾患」の項目へのリンク対応づけがあると、検索者は「消化」に関連する図書群をさらに細かい観点から、また関連する別の図書も検索できる。

このようなリンクを実現するためのガイダンスデータはリンク元の分類体系の項目と、リンク先の分類体系の項目の対である。このガイダンスデータは以下の手順で自動的に作成される。与えられた分類体系の各概念項目間に関連度を設定した関連度マップを作成する。そして、各分類体系の項目毎に評価を行い、最も評価の高い場所を適切なものとし関連付ける。

関連度マップの作成には、分類体系で使用されている概念を表す単語が対象図書群全体に出現する頻度を用いて算出する。これは前回報告[1]した分類体系の階層間の関連度定義と同方式である。ある図書中に2つの単語が出現した場合、何らかの関連があると考え、共出現頻度を算出し各単語の総出現頻度に対する比率で関連度を定義する。これを全ての分類基準の項目に対して算出し単語関連度マップを作成する。

3. 試作システム

以上の考えに基づき、ワークステーション上でガイダンスデータ作成を行い、ネットワーク経由でパソコンから図書検索するシステムを試作した。試作システムに試験データを投入し実験、評価を行う予定である。

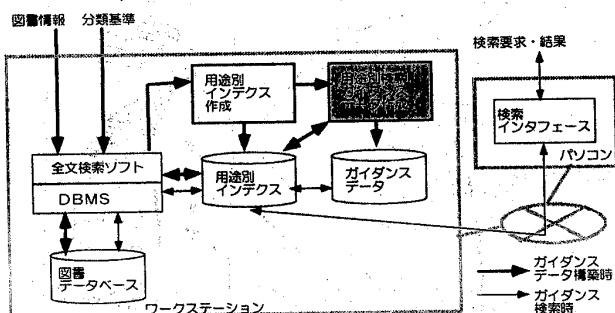


図2 システム構成図

実験は医学分野を対象とし、図書の代わりに技術文献を用いて実験を行なう。

・使用文献データ: JICST科学技術文献速報CD-RO

Mライフサイエンス編の文献(約9万件)

・使用分類体系: 日本10進分類(NDC)、米国医学図書館分類法(NLM)、看護大全(一般向医学書)目次
・頻度評価対象: 各文献の200文字程度の抄録部分

評価項目を明らかにするため、この方式における問題点を机上評価した。

(1) 計算量

関連度マップを作成する際の計算量は、分類体系に含まれている単語を基に各文献毎に走査、相互の関連度を計算するため、対象文献数および分類体系に出現する総単語数の2乗に比例する。対象となるデータの規模により計算量が爆発する可能性がある。現状の試験データにおける計算量を1とすると、同規模の新規の分類体系が1つ増える場合で、計算量は4になる。

関連度の定義、計算アルゴリズムの見直し、また、分類体系に変更が生じた場合の関連度マップの修正は、最初から再計算を行わず差分情報による修正方法の検討が必要である。

(2) リンク先決定の考え方

分類体系のある項目において検索できる図書と、ある程度同じ図書が検索できる項目を最良と考えているため、評価方法によっては分類体系の階層の上位を指向しがちになる。約20%程度が第2階層目(全体を15分野程度に分類している層)へのリンクを最良と評価している。結果として、大きすぎる概念を提示することになり、ガイダンスとしての意味が薄れ、関連の深いものとはいえなくなる。上位の階層を選択する方向を抑制する別の評価式の検討や、関連度評価が低いものはリンクの対象としないようにするための閾値の検討などの必要がある。

4. おわりに

我々は、用途別検索のためにガイダンス機能の手法を検討し、試作を行った。机上評価では、この方式では、計算量の問題、関連付けの際の考え方の問題などが、明らかになっている。今後、実験、評価を行い、計算量の削減、用途別インデックスを生かした新しい検索手法の検討などを課題として進めていく。

謝辞

本研究開発は、通商産業省「次世代電子図書館システム研究開発事業」に参画し、実施したものです。本事業は、国からの委託を受けた情報処理振興事業協会(IPA)より、(財)日本情報処理開発協会(JIPDEC)が再委託され実施しているものです。

参考文献

[1] 嶺岸、田中: “用途別統合検索における図書検索”, 情報処理学会第56回全国大会 4Y-4