

概念マップを用いた図書検索支援システム

4 Y - 5

水口政雄* 水野舜**

*金沢工業大学大学院 工学研究科 情報工学専攻

**金沢工業大学 人間情報経営系

1. はじめに

従来の図書検索のシステムでは、入力されたキーワードと書名とのマッチングやツリー構造の概念を辿る方法で必要な図書を限定していた。しかし、キーワード検索では類義語や関連語の過不足が再現率・適合率に大きく影響する。また、複雑な階層構造で表される概念を効率よく探索するためには、データベース全体に対する概略的な知識を持つことが要求される。

これらの問題を解消する方法として、本研究ではキーワードと概念集合から自己組織化処理によって概念同士の位相関係を表した2次元のマップ(以下、概念マップと呼ぶ)を形成し、利用者との対話の中で提示する事でデータベース全体の概観と要求の指す位置を与える。また、その作成過程で抽出された単語を関連語として提供することで利用者のキーワード入力の負担を軽減する。

2. 検索支援システムの構成

提案するシステムの構成を図1に示す。

本システムは事前に概念マップを作成しておくマップ作成部と、実際の検索時に動作するインタフェース部から成り立っている。

マップ作成部ではあるまとまった概念ごとに図書データを取り出し、単語の出現頻度により特徴付けを行った後、Kohonen の自己組織化マップ処理[1]によって概念マップを形成する。

実際の検索に際してインタフェース部では検索結果の図書集合とともに現在のユーザの要求が指す位置と関連語情報を返す。

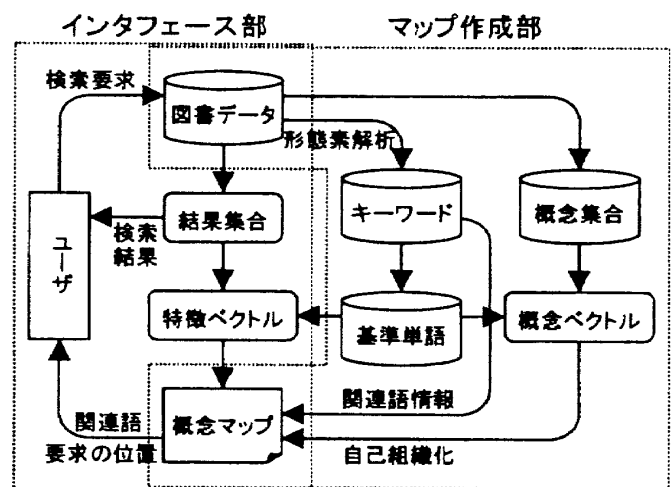


図1. システム構成

3. 概念マップの作成

3.1. キーワード抽出

すべての図書データからキーワード検索の対象項目である書名及び副書名を抜き出し、形態素解析を行う。解析にあたって、キーワードの誤分割を押さえるため、名詞の連続性に着目した成句検出[2]を並行して行う。切り出された単語の中からキーワードになりうる品詞として、名詞(普通名詞、サ変名詞、時相名詞、固有名詞)と形容詞をその出現回数とともにキーワードテーブルに登録する。

3.2. 基準単語による概念のベクトル化

利用者が導き出そうとする概念と入力すべきキーワードの間には何らかの相関があると思われる。そこで、特徴となるような単語(基準単語)の頻度情報を用いて概念集合をパターン化する。具体的には、基準単

語の種類を次元とし、その出現頻度が各要素に比例する様なベクトルで表し、これを概念ベクトルと定義する。なお、概念集合として、ほとんどの図書に付けられているNDC(日本十進分類)をもとに検索された図書の集合を利用する。

3.3. 概念マップの作成

Kohonenによって提案された自己組織化マップは、2層のニューラルネットワークに出力ユニット同士の位置関係を持たせ、近傍学習を行う事により、入力パターン間の自然な位相関係を2次元平面上に表す事ができる。つまり、通常の競合学習によるクラスタリングに加えて特徴の似ているパターンがマップ上の近くの点にまとめられるという性質があり、テキストデータベースにおける文書の自動分類[3]や曖昧検索[4]などに応用する試みがなされている。

本システムでは、概念ベクトルから自己組織化によって概念の特徴マップを作成し、出力層の各ユニット毎にNDCの件名標目表から得た概念見出し語と関連語を付加する。これにより、情報散策の手がかりを与えると同時に、膨大なキーワードから各カテゴリの書名に実在する関連語だけを与える事が可能になる。

4. 検索作業の支援機構

4.1. データベースの概観と位置情報の提示

一度キーワードで検索した際、基準単語を用いて検索結果の集合からその特徴ベクトルを作成し、検索結果と共に概念マップ上のクラスタリングされた位置を視覚的に返す。また、各出力ユニットの重みとの誤差から、近い分野の順にカテゴリを導き出すことも可能である。

4.2. 関連語情報の表示

マップ上のユニットには、その概念に属する関連語情報が付加されている。それらは出現頻度順に並べて表示でき、利用者のキーワード入力を支援する。

5. 終わりに

概念集合の特徴ベクトルから生成した概念マップを用いて図書検索作業を支援するシステムについて述べた。本システムにより得られる利点として次の様なことが挙げられる。

- ・データベース全体の概観を与える事が可能になり、従来のシステムでは不可能だった情報散策がある程度可能になる。
- ・概念同士の類似度もマップ上の位置という形で表せるため、概念間の移動に伴う利用者の位置記憶の負担を軽減できる。
- ・実際に書名の中で用いられている膨大なキーワードの中から、利用者の求める概念に合った単語のみを関連語として提供できる。
- ・概念同士のリンク関係や関連語情報を静的に決定することなく、実在するデータから全機械式で作成できる。

今後はデータの更新に柔軟に対応していくために、差分学習の機構等を考慮していく必要がある。

参考文献

- [1] Kohonen, T., The Self-Organizing Map, Proceedings Of The IEEE, Vol.78, No.9, 1990
- [2] 神林隆, 清水奨, 佐藤進也, Paul Francis, 「インターネット情報検索に適したキーワード抽出」, 情報処理学会研究報告 97-NL-118, 1997
- [3] 津高新一郎, 「自己組織化マップを用いたテキスト分類の試み」, 情報処理学会第46回全国大会, 分冊4, pp. 187-188, 1993
- [4] 銭晴, 史欣, 田中克己, 「自己組織化マップと語彙索引を用いたデータベースの抽象化機構」, 情報処理学会研究報告 94-DB-99, 1994