

データベース知的検索システム

4 K-7 ファジィ関係を用いたあいまい表現

及川 和彦* 岡村 博之* 平松 斉**
 * 三菱電機(株) ** 三菱電機東部コンピュータシステム(株)

1. はじめに

近年、商用データベースの数、利用率は、年々大きな伸びを示しているが、商用データベースに対するユーザの意識においては検索したい対象になかなかたどりつけないという機能的な不満が多く聞かれる。更に、各企業の技術情報に関しては、利用の増大と多様化、情報量の増大と情報媒体の多様化といった状況の変化と生産性向上という面から社内データベースを構築することが、最近の傾向となっている。このような状況の中で、情報検索あるいは情報管理の重要性が認識されてきている。本論文では、上記背景を踏まえて開発されたデータベース知的検索システムについて述べる。

2. データベース知的検索システム

2.1 システム概要

本システムは、社内データベースにおける技術文書のキーワードによる検索を効率的に行うことを目的としたシステムであり、図1のシステム構成図に示すように2つのサブシステムからなる。

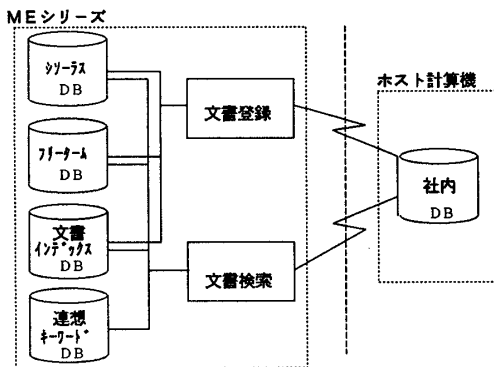


図1 システム構成図

2.1.1 文書登録サブシステム

ユーザの技術文書の登録作業を支援するサブシステムである。登録の際には、技術文書に関する情報（文書名、作成者、作成日付、検索キーワード etc）をユーザに設定してもらい、内部データベースであるソース、フリータム、文書インデックスに展開する。

本サブシステムの目的は、文書検索サブシステムが行う知的検索のための情報の作りこみである。

2.1.2 文書検索サブシステム

ユーザの意図する社内データベース内の技術文書の検索をより効率的に支援するサブシステムである。

本サブシステムは、ユーザの概念に最も関係の深いソース/フリータムキーワードを導き出し、更に、

それらの組み合わせである検索論理式をよりユーザの意図に近い形で自動生成し、検索すべき技術文書の含まれるグループの絞り込みを行う。

2.2 システムの特徴

①プラン管理

システム全体の処理シーケンスをシステムプランとして保持しており、その時の状態の遷移と照合しながら処理を進めていく。

②連想キーワード検索

ユーザの概念に最も関係の深い技術文書に結び付くソース/フリータムキーワードを例示する。ユーザ概念と実際のキーワードとの関係付けは連想キーワードデータベースで行っている。

③関連語検索

ソース/フリータムキーワード間の関係（同一文書内の共出現頻度）を用い、最も関係の深い他のソース/フリータムを導き出す。ソース/フリータム間の関係付けは、ソースデータベース及び、フリータムデータベースで行っている。

④簡易学習

- ・検索の際に全くの未知語を入力された場合、システムが把握できなかった概念として一時保存しておき、後の検索時に入力されたシステムが把握できる概念との関係付けを行う。
- ・文書登録サブシステムで技術文書を登録する際に設定される検索キーワード間の関係を更新する。
- ・最終的に文書を社内データベースから取り出した（文書引用）時に、検索条件として設定されたキーワード間の関係を更新する。

⑤論理式自動生成ルール

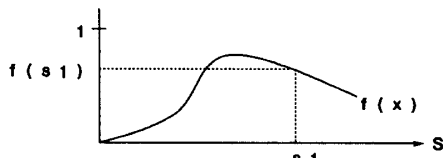
検索条件として設定されたソース/フリータムキーワード間の関係を利用して、設定されたより多くのソース/フリータムキーワードが条件として効果が現われるように検索論理式を組み立てる方法をルールとして実現している。

3. ファジィ関係によるインプリメント

3.1 同一文書内共出現頻度

同一文書内共出現頻度とは、SaltionらのSMARTシステムでの研究^{[1][2]}が前提となっており、ソース/フリータムキーワードが技術文書の登録の際に共に付加情報である検索キーワードとして設定されている頻度（回数）のことである。つまり、A、Bというソース/フリータムキーワードが4つの技術文書に検索キーワードとして登録されている場合、A、B間の共出現頻度は4となる。本システムでは、この共出現頻度をファジィ関係で表現し、ソース/フリータムキーワード間の類似性、同一目的性といったあいまいさの表現を実現している。

共出現頻度は、文書登録サブシステムによって登録をくり返すことで更新されていく。(簡易学習)



$f(x)$: あるシソーラス/フリータームキーワードと他のシソーラス/フリーターム間の共出現頻度による関係のあいまいさを表わすメンバシップ関数
 S : あるシソーラス/フリータームキーワードと関係を持つシソーラス/フリータームキーワードの全体集合
 $s1$: シソーラス/フリータームキーワード
 $f(s1)$: あるシソーラス/フリータームキーワードと $s1$ のあいまいさを表わす値

図2 同一文書内共出現頻度の実現方式

3.2 検索頻度

検索頻度とは、本システムでは連想キーワード検索と実際の社内データベースから技術文書を取り出した(文書引用)時に適用される。

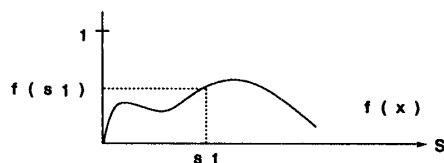
①連想キーワード検索における検索頻度

連想キーワード検索時のユーザ概念とシソーラス/フリータームキーワード間の関係の度合いを検索頻度で表わしている。検索頻度は、連想キーワード検索でユーザ概念からシソーラス/フリータームキーワードの例示が行われ、その中から検索条件としてユーザが選択したタイミングで更新される。

②文書引用における検索頻度

文書引用するときに検索条件としてユーザが選択したシソーラス/フリータームキーワード間の関係の度合いを検索頻度で表わしている。同じ検索条件で文書引用を繰り返すことによって、検索条件であるシソーラス/フリータームキーワード間の関係付けは強化され、次の文書検索における論理式自動生成ルールの結果に影響を与えていく。

本システムでは、同一文書内共出現頻度と同じく検索頻度についてもファジィ関係で表現し、あいまいさの表現を実現している。



$f(x)$: あるシソーラス/フリータームキーワードと他のシソーラス/フリーターム間の検索頻度による関係のあいまいさを表わすメンバシップ関数
 S : あるシソーラス/フリータームキーワードと関係を持つシソーラス/フリータームキーワードの全体集合
 $s1$: シソーラス/フリータームキーワード
 $f(s1)$: あるシソーラス/フリータームキーワードと $s1$ のあいまいさを表わす値

図3 検索頻度の実現方式

3.3 検索論理式自動生成ルール

検索論理式自動生成ルールは、上記同一文書内共出現頻度と検索頻度を利用し、できるだけユーザの意図を反映する形でシソーラス/フリータームキーワードの組み

合わせである検索論理式を生成する。

代表的なルールの例を以下に示す。

(例1)

選択されたシソーラス/フリータームキーワードの内、最初に選ばれたものは検索条件の最上位のプライオリティにする。

(例2)

選択されたシソーラス/フリータームキーワード間に共出現頻度あるいは、検索頻度によるファジィ関係があるものは、アンド結合でグルーピングする。

4. 評価

本システムは、単にルールを用いたエキスパートシステムとしての機能だけでなく、各技術文書に付加される検索キーワード間の関係(あいまいさの表現)にファジィ関係を導入することにより、より柔軟で効率的な文書検索を可能にした。更に、ファジィ理論の応用例の多くは、主観を扱っているということから、そのメンバシップ関数の必然性、有効性というものがあいまいであり、説得力にかけてしまう面があった。この点では、Salt onらの研究⁽¹⁾⁽²⁾に見られる共出現頻度という1つの基準の元にあいまいさを表現していることから、メンバシップ関数自体の不透明さを解消する1つの方法を見出せたと考えている。但し、本システムにおいても、以下のような大きな課題が残されていると認識している。

①シソーラス/フリータームキーワード間のあいまいさの妥当性は、文書登録時のユーザの設定する検索キーワードに大きく左右されてしまう。この点については、文書登録サブシステムの強化として、キーワード自動抽出機能を考える必要がある。

②本システムは、シソーラス/フリータームキーワード間の関係の表現にファジィ関係を応用したものである。今後は、ファジィデータベース言語の研究と結び付けることが必要と考えている。

5. おわりに

データベース知的検索システムについて、主に、シソーラス/フリータームキーワード間の関係のあいまいさの実現について述べた。

現在、本システムは実用システムとして、フィールドテスト中である。今後更に、実用化に向けて改良を加えていく予定である。

《参考文献》

- [1] Salt on G.(ed.).(1971): The SMART Retrieval System.Experiments in Automatic Document Processing.Prentice-Hall,Englewood Cliffs,NJ
- [2] 三宅輝久、宮本定明、中山和彦:『ファジィ集合に基づくデータベースの索引語関係の生成と情報検索への応用』第4回ファジィシステムシンポジウム講演論文集
- [3] 小林健三、佐々木道夫、梁田和彦他:『文献データベース知的検索システム』三菱電機技報1988.10