

3R-1

ファジィ情報検索のための
FAIRS-I 情報検索システムの拡張機能

三宅輝久、宮本定明、中山和彦
(筑波大学)

1. はじめに

近年、情報検索が盛んに行われるようになるに従い、検索を効率良く行うための手法が求められてきた。その一つとして、ファジィ集合理論を応用して検索質問に対する結果の文献に対してメンバシップの値を付与し、その値により結果の文献の評価を行うことができるファジィ情報検索がある。本稿では、既存の情報検索システムであるFAIRCOM FAIRS-Iに対してファジィ情報検索機能を組み込み、通常の実験と同様にファジィ情報検索を可能とするシステムを試作したので報告する。

2. ファジィ情報検索

ここで拡張したファジィ情報検索機能は、次の二つのステップに分けることができる。

(1)、ファジィ化シソーラスによるキーワードの展開

検索質問 q 中のキーワード w はファジィ化されたシソーラス F により、その関連語 (Relative Term) v_1, \dots, v_n のファジィ集合 q' に展開される。(図1)

$$q' = \sum F(w, v_i) / v_i$$

(2)、ファジィ化索引による検索

展開された個々のキーワード v_i についてファジィ化索引 I により検索を行い、検索結果の文献集合全ての和集合をとる。検索結果の文献のファジィ集合 r については、 q' と I との間のファジィ関係の合成演算によって求められる。(図2)

$$r = \sum_i \max \min [F(w, v_i), I(v_i, d)] / d$$

3. システムの構成

ファジィ情報検索機能は、一つのユーザルーチンとしてFAIRS-Iに組み込まれ、そのインターフェースルーチンであるXFLINKを用いてFAIRS-Iの検索機能を利用している。利用者は、通常の実験を行う途中でファジィ情報検索機能呼び出し、終了後はまた通常の実験に戻ることができる。ファジィ化シソーラスは一つのデータベースとして作成され、FAIRS-Iの機能を用いてアクセスされる。また、ファジィ化索引は、実際には、通常の実験とメンバシップ値に分けられている。索引の部分はFAIRS-Iのインバーテッドファイルを利用している。また、メンバシップ値はデータベース中の個々の文献レコード中に一つの項目データとして格納されている。(図3)

4. ファジィ情報検索用のコマンド

ファジィ情報検索用に用意されたコマンドを表1に示す。各コマンドは基本的にはFAIRS-Iの実験コマンドに

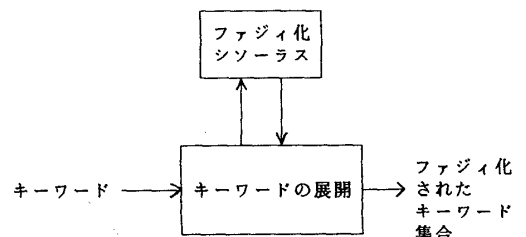


図1 ファジィ化シソーラスを用いたキーワードの展開

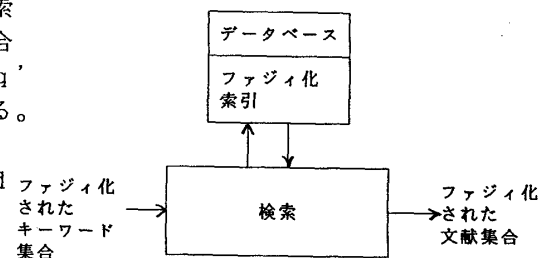


図2 ファジィ化索引を用いた検索

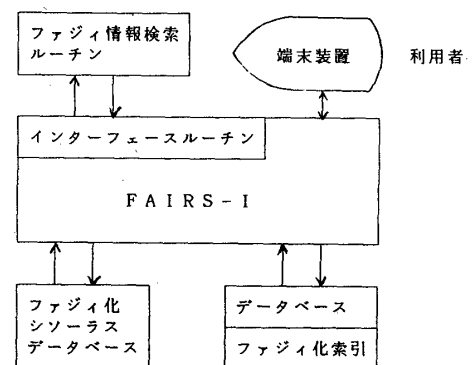


図3 システムの構成

対応している。検索コマンド (FSEARCH) では、ファジィ化シソーラスによるキーワードの展開とファジィ化索引による検索の二つのステップを連続して行う。表示コマンド (OUTPUT) では、検索結果の文献集合をメンバシップ値により四つのグループに分け、オペランドの指定がない場合には最もメンバシップ値の高いグループが出力されるようにしている。ファジィ情報検索機能を利用した例を図4に示した。

5. 問題点

現在のファジィ情報検索機能には、以下の問題点が残されている。

(1)、ファジィ化索引の作成

ファジィ化シソーラスについては、索引語の出現頻度に基づいて自動的に生成する方法について既に報告した*2。しかしながら、ファジィ化索引については有効な作成方法が示されてはいない。(今回使用したファジィ化索引はシソーラスと同様索引語の出現頻度に基づいて作成した。)

(2)、検索時間

従来の検索と比較して、メンバシップ値の低い多数の文献を合わせて検索することになるため、検索時間が長くなる。また、ファジィ化索引を実現する有効な索引手法がないため、処理時間が余計に必要である。

6. おわりに

今回報告したファジィ情報検索機能は、既存のシステムに容易に組み込むことが可能であり、大規模な文献データベースに対しても適用することが可能である。現在はいくつかの問題点が残されているが、それらについては今後検討して行きたい。

〔参考文献〕

- 1、三宅、宮本、中山、「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」、1988、第4回ファジィシステムシンポジウム
- 2、Miyamoto, Miyake and Nakayama, Generation of a fuzzy pseudothesaurus for information retrieval based on cooccurrences and fuzzy set operations, IBBE TRANS., Syst., Man and Cybern., Vol. SMC-13, No. 1 (1983)

表1 ファジィ情報検索用のコマンド

1. FINIT.....Start fuzzy information retrieval
2. FBROWSE.....Browsing terms in the fuzzy thesaurus
3. FSEARCH.....Retrieve database
4. FSAVE.....Save the current set of documents
5. FCLBAR.....Delete a saved set
6. FCOMBINE....Perform a fuzzy logical operation to saved sets
7. FOUTPUT.....Display documents in current or saved set
8. FLOAD.....Load a saved set to the current set
9. FSHOW.....Display the layers of the current set
10. FVALUE.....Display the membership values of the current set
11. FEND.....End fuzzy information retrieval

通常の検索

ファジィ情報検索の開始

```
> FINIT
*** FUZZY ASSOCIATION SEARCH STARTED ***
> FSEARCH 'FUZZY SET THEORY'
+ACT0011 ACCOUNT INFORMATION. < DB :
*** RESULT : TERM(S) = 8
              : LAYER 0 2 DOCUMENT(S)
              : LAYER 1 707 DOCUMENT(S)
              : LAYER 2 1168 DOCUMENT(S)
              : LAYER 3 0 DOCUMENT(S)
              : TOTAL 1877 DOCUMENT(S)
> FSAVE SET(%FUZZY)
*** SET(%FUZZY) SAVED
> FSEARCH 'MACHINE READABLE MATERIALS'
+ACT0011 ACCOUNT INFORMATION. < DB :
*** RESULT : TERM(S) = 5
              : LAYER 0 1 DOCUMENT(S)
              : LAYER 1 501 DOCUMENT(S)
              : LAYER 2 748 DOCUMENT(S)
              : LAYER 3 1 DOCUMENT(S)
              : TOTAL 1251 DOCUMENT(S)
> FSAVE SET(%MRM)
*** SET(%MRM) SAVED
>
> FCOMBINE %FUZZY AND %MRM
              : LAYER 0 0 DOCUMENT(S)
              : LAYER 1 48 DOCUMENT(S)
              : LAYER 2 25 DOCUMENT(S)
              : LAYER 3 26 DOCUMENT(S)
              : TOTAL 99 DOCUMENT(S)
> FOUTPUT 1
*** DOCUMENT(LAYER 1) OUTPUT ***
```

```
# 1 OF 48 / S= 0.00549
IDNO 87/998
A SINITSYN, A.N.;SHAVANICHE
CA VINITI (ALL-UNION INSTITI
INFORMATION), USSR
T USING GROSS-FORMULAE OF
COMPOUND'S STRUCTURE
TO PRIMENENIE BRUTTO-FORMI
EGO STRUKTURY
J NAUCHNO-TEKHNICHESKAYA
B SERIES 2 (2) 1986, 15-
YR 1986
LG RUSSIAN
C ZJM(547:541.6)
CH ORGANIC CHEMISTRY. CHEM
VINITI (ALL-UNION INSTITI
INFORMATION)
D U.S.S.R.;PROGRAMS (COMPUT
CHEMISTRY;CHEMISTRY;TECHN.
STORAGE AND RETRIEVAL;INFO
INDEXING;COMPUTERISED INFO
ISED SUBJECT INDEXING;SUBJ
SYSTEMS DEVELOPED IN THE U
STRUCTURE-OF-ORGANIC COMPO
ON SPECTRAL INFORMATION P
FRAGMENTS. THE MOST UNIVF
BOND BETWEEN ATOMS WITHO
DATA ARE IN THE FORM OF
COMPOUND BONDS DEVELOPF
GROSS-FORMULAE ARE THE
EXCLUDING ALL THOSE W
INFORMATION AND THE '
'MODEL' FOR THE COMF
```

ファジィ情報検索の終了

通常の検索

図4 ファジィ情報検索の例