

6M-2

レファレンスデータベース利用のための
知識表現について

伊藤 秀昭 米田 順美 浜中 栄治
財団法人 日本情報処理開発協会 開発研究室

1. はじめに

近年、知識型システムが様々な分野において開発が進められている。その一つの応用分野に、データベース検索および情報検索のための知識型システムを構築する試みがある〔MCCU, ZARR〕。現在、レファレンスデータベースとしてデータベース台帳総覧〔TUUS〕に記載されている内容の一部を再構成し(クリアリングデータベース〔KIKI〕と呼んでいる)、これを用い利用者の目的に沿うデータベース(DB)を選択するためのシステムの開発が進められている。現時点におけるクリアリングDBシステム(CDBS)は、利用者が問い合わせに用いることの可能な分野を表すためのキーワードが固定されているため利用者固有のキーワードを検索に利用することが困難である、検索が文字列を対象としているため融通性に欠けるなどの課題を有している。現在、我々はこのような課題を解決するために、CDBSが取り扱う範囲を分野(現時点では、工学関係)に限り限定し、NIRS(Network-based Information Retrieval System)と呼ぶ知識型システムの開発を進めている。本稿においてはNIRSの知識ベースの構造について述べる。

2. システム構成

NIRSの構成を図1に示す。DBOsは、DBO(CDBSが有する個々のDB)より成る。知識ベースは、DBO 検索のためのルールを記述するルール部、および検索のためのキーワードを記述するためのネットワーク部より成る。ルールはIF条件部THEN結論部の形式により表されるプロダクションルールであり、ネットワークは学問分野などを表す名詞をノードとし、ノード間を結合するリンクから成る。ここでは、ノードを記述子、ネットワークを記述子グラフ(DG)と呼ぶことにする。このように構成した理由は、検索のための判断がルールとして記述が容易であり、検索に利用される記述子間の関係(例えば、工学—情報工学—コンピュータ)が直観的な理解が容易である、と考えたからである。このように表現された知識を解釈し、問い合わせに答えるメカニズムが、推論機構である。これは、ルールとDGに基づきDBO および問い合わせに現れる記述子と関連する記述子を得ることが可能である。推論機構により関連づけられた記述子をDBO の概要(NIRS においてはテキストではなく、元のテキストより得られた名詞となっている。)に含むDBO を得ることができる(図2)。ただし、現段階ではDBOsはNIRSの構成要素となっているが、これを分離しCDBSとの結合を図ることは、今後の課題の一つである。

3. 知識ベースの構造

3. 1 記述子グラフの記述

DGは、主題、一般語および固有名詞の三種の記述子のク



図1 NIRSの構成

ラスより成る。それらは、分野における学問領域(機械工学、電子工学)を表す名詞が、手法、理論などの比較的多くの分野においても出現すると思われる名詞が、学会名称、関連ファイルなどの固有名詞が、それぞれ記述子として属している。このように記述子を分類した理由は以下のようなことである。DBO に含まれる概要では、各々のDBがどのような分野の文献、データなどを収録しているのか、そしてその特徴としていかなる観点より収録したDBであるのかという話題が中心に記述されている。そこでは名詞がDBOの特徴を表すために中心的な役割を果たしていると考えられる。したがって、記述子を分類することによりDGの構造が明確となると考えられる。概要に現れる名詞はいくつかのクラスに分類することが可能であり、現段階では上記三種により十分であると考えている。

二つの記述子を結合するためのリンクは、次の五種が用意されている。それぞれには0から1までの数値による重みを付加することが可能であり、結合される二つのノード、重み間の関係は三項述語のように書かれる。ここで、A, Bは記述子、wは重みである。このとき、重みは、記述子間の関連性を数値として表していると考えている。

i. GEN(A, B, w): 記述子Aの上位語はB。上位語を複数定義することも可能である。

ii. SPE(A, B, w): AはBの下位語。GENとSPEは記述子間の階層構造を設定するリンクであり、GEN(B, A, w1)はSPE(A, B, w2)と同様である。ただし、重みw1, w2は異なる場合がある。これは、利用者により関係が明示されないならば、暗黙に設定されるリンクに基づく推論は行われなことを示している。もしSPE(A, B, w)と宣言されたのみならば、リンクの解釈はGEN(B, A, 0)と同様である。

iii. SYN(A, B, w): AおよびBは同義語。例えば、SYN(コンピュータ, 電子計算機, 1.0)。ただし、リンクは双方向性を有している。

iv. ASS(A, B, w): BはAより連想される記述子。例えば、ASS(機械工学, 設備, 0.4)。

v. SEA(A, B, w): See alsoの略であり、関連のある分野に対する参照を表すためのリンク。ASS(A, B, w)と同様に記述子間の関連性を表しているが、A, Bは共に分野を表す記述子である。

3. 2ルールの記述

(1)利用者定義の記述子記述。

問い合わせの記述において利用する記述子を利用者自身が設定し、それを知識ベースに格納することにより、利用者固有の記述子によるDBO 検索を行うことを実現しようとするためのルールである。この種のルールは、利用者定義の記述子導入による融通性の向上および検索手続きの記述を明快にすることを可能とすると考えている。

例えば、次のような問い合わせを考える。

"SUBJ(機械工学) and AREA(EC) and SOUR(特許)"

この問い合わせは、"EC における機械の特許情報を含むDBは何か"、ということを表していることにする。ここで、SUBJ (機械工学) はDGにおいて定義されている分野が機械工学であり、AREAおよびSOURはDBO 記述において文献収録地域がEC、情報源が特許であることをそれぞれ示している。ここで、もしECが記述子として設定されていないとするならば、利用者は、情報収録地域をEC加盟国個々について問い合わせを記述しなければならない。このとき、AREA (EC) がルールとして、次のように定義されているとする。

```
IF AREA (EC) THEN AREA (イギリス) or
                        AREA (フランス) or
                        ..... or
                        PROP (ヨーロッパ)
```

ここで、PROPはDGにおけるクラス固有名詞を表し、ヨーロッパは固有名詞の一つとして設定されていることにする。この種のルールは、問い合わせ文に対する書替え規則のように利用される。したがって、先の問い合わせは、AREA (EC) がルールの結論部により書き替えられる。このように、AREA (EC) を規則の右辺により書き替えるというルールを用いることにより、各国毎に問い合わせを記述するという操作の軽減化を図っていると考えられる。

(2) 分野とDBO の結合。

ある特定の分野を指定することにより、指定された分野に関する文献、データを格納するDBO が決定される場合が多々ある。これは、あるDBが分野に関して特徴を有している場合である。実行効率の観点から、DGを用いた推論に基づく検索は低いと考えられる。したがって、特徴のある分野ならば、直接的にDBO を求めるメカニズムをシステムが提供していることが適切である。このような知識は、判断規則の一種であるため、ルールとして容易に記述することが可能であると思われる。この種のルールは、"IF SUBJ(A) THEN DBO, w." の形式により書かれ、w は個々のルールに設定された確からしさとして解釈されるとする。

NIRSでは、この種のルールを簡潔に記述するために、次のように記述することを許している。

```
r1: IF r2="true" and A THEN DBOi , w1.
r2: IF C and D THEN DBOi , w2.
```

ここで、ルールr1は、ルールr2が真でありかつ条件A が成立するならば、DBO_i であるという結論は重みがw1であるということを表している。これは、特徴のある項目が付加されることにより、あるDBO であるという確信が高まる、また逆にある条件が成立しないために確信が低下するというような規則を容易に記述するために導入されている。

4. まとめ

本稿では、NIRSの知識表現の概略について述べた。推論メカニズム、問い合わせ処理および実現などについては機会を別に述べる予定である。DN構築では、現在インタフェースが充実されていないためシステムの操作性は劣っている。例えば、DGの管理、階層およびリンクのパスなどに対する見通しを持っての構築方法などである。今後、利用を通じての評価、インタフェースの拡充および実行の高効率化などが課題であり、検討を進める予定である。

謝辞 クリアリングデータベースシステムは、(財)データベース振興センターの協力の下で開発されている、利用を許して頂いたデータベース振興センターに深謝致します。

参考文献

(ITO) 伊藤, 滝沢, 上野 : フレーム型データ構造における論理的構造について, 信学会, AI86-19, 1986
 (KIKI) 機械システム振興協会: 高度データベースシステムの開発に関するフィージビリティスタディ報告書, 1986
 (MCCU) McCune, B.P., Tong, R.M., et al: RUBRIC: A System for Rule-based Information Retrieval, IEEE Tran. SE, Vol. SE-11, No.9, 1985
 (SALT) Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval, McGraw-Hill, 1983
 (TUUS) 通産省: データベース台帳総覧
 (ZARR) Zarri, G.P.: Expert Systems and Information Retrieval: an Experiment in the Domain of Biographical Data Management, Developments in Expert Systems, ed., Coombs, M.J., Academic, 1984

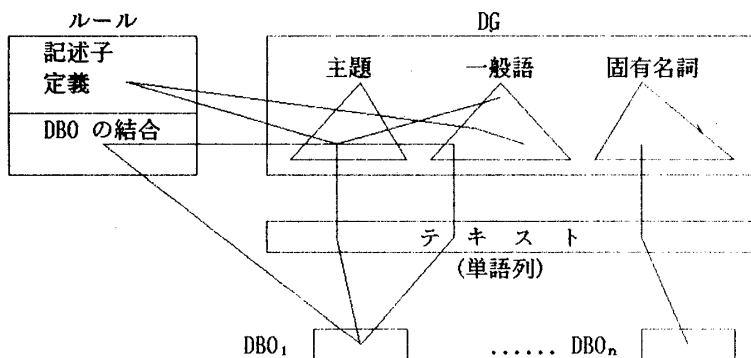


図2 知識ベースの概略

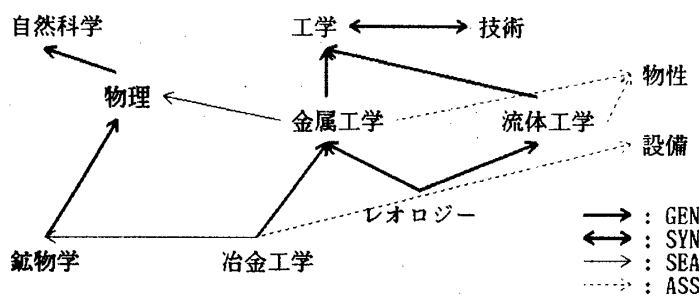


図3 DG記述の一例