知的情報検索システムＩＲＩＳにおけるキーワード検索式生成方式

秋山　幸司
（富士通株式会社）

1．はじめに

知的情報検索システムＩＲＩＳは，簡単な日本語で入力された検索要求について，その回答となる意味内容を持つテキスト群をテキストベースから検索することをめざす実験的システムである〔杉山86〕．現在，ＩＲＩＳでは，数百件のテキストベースを対象としたプロトタイプ版の評価を終了し，数千件のテキストベースを対象とした拡張版を試作中である．本稿は，ＩＲＩＳ拡張版の検索式生成部ＡＲＥＸ（Advanced Retrieval EXpert）における拡張・改良を述べるものである．

2．ＡＲＥＸの概要

ＡＲＥＸは，検索要求を表す日本語質問文を解析して得られた意味構造から，人間のキーワード検索専門家の持つ技能を用いて，要求の回答となりうるテキスト群を検索するためのキーワード検索式を生成し実行するエキスパートシステムである．ＡＲＥＸの入力はＩＲＩＳの質問文解析部が生成した意味構造である．ＡＲＥＸは，専門家が情報検索を行う際に検索を一歩進めるために使う基本操作（検索戦術〔bates 79〕），検索戦術の適用順序を決める検索戦略，各検索戦術の適用可能性を判定する戦術適用規則を知識として持つ．検索戦略が示唆した検索戦術を戦術適用規則でチェックしながら用いて入力意味構造を変形展開することで検索式を生成する．この検索式を実行してテキストベースを検索することにより，検索結果を得る．紙面の制限上，ＡＲＥＸの基本的枠組や本稿で使う用語の意味については〔秋山 87〕を参照されたい．

3．プロトタイプ版ＡＲＥＸの問題点

ＩＲＩＳプロトタイプ版について，ＡＲＥＸに関係する事項を評価した結果，以下に示す問題点が判明した．

(1)分野モデルの不備

ＩＲＩＳプロトタイプ版の分野モデルでは，述語的概念と名詞的概念との関係（いわゆる深層格）を中心に記述し，名詞的概念相互の関係はほとんど記述せず，名詞相互の関係も解析しなかった．

この手法でも，ＡＲＥＸによる検索結果は実用レベルであったが，戦術適用規則の記述の容易性・可読性に難があることが判明した．特に，検索に不要なキーワードを削除する縮約戦術においては，名詞概念間の関係を用いない限り精密な削除は不可能であり，無用な削除による処理時間の増大および適合率の悪化が生じていた．

(2)固定的戦略

プロトタイプ版ＡＲＥＸでは，対象分野の典型的な内容の検索を目標にした．そこで，検索すべき内容に応じた動的な検索戦略の生成は行わず，検索専門家が使う最も基本的な戦略一つに固定した．しかし，このような固定的検索戦略では，無駄な推論を行う場合や最適な検索式が生成できない場合が生じた．

(3)推論（戦術適用）の停止条件

ＡＲＥＸの検索戦略は，生成される検索式の厳密さを戦術の適用に従って徐々に緩めるものとしている．従って，再現率や適合率などの各種検索効率が総合的に最適となるところで推論（戦術適用）を停止しなければならない．プロトタイプ版ＡＲＥＸでは，ユーザが指定する検索結果件数期待値を閾値として推論を停止していた．この閾値は，テキストベース中の検索要求を満たすテキスト件数と密接な関係にあり，値の推測を誤ると検索漏れや大量のゴミが発生したため，結果件数に依存しない推論制御の必要性が判明した．

4．改善手法

前節で述べた問題点を改善するため，ＩＲＩＳの分野モデルおよびＡＲＥＸの推論機構を以下に示すように拡張・改良した．

(1)名詞的概念間の関係の定義

分野モデルの名詞的概念について，実体間や実体－属性間の関係を中心に名詞的概念間の関係および意味的制約を規定した．また，連続した名詞列についても格助詞「の」で連体修飾された場合と同様にそれら相互の関係を顕在化するように，構文意味解析部を変更した．これを受けて，ＡＲＥＸにおける戦術適用規則をこれらの関係を用いた記述に書き換え，また，その一部をより精密化した．

(2)意味構造間の意味的距離指標の導入

ＡＲＥＸの本来の目的は，検索要求が示す内容に類似の意味内容を持つテキストを件数にかかわらず全て検索することである．類似性を一般的に定義することは不可能だと言えるが，ＡＲＥＸでは，ある戦術をある条件下で適用すればどの程度類似性が低下するかという相対的な値を与える知識が記述できればよい．この知識は戦術適用規則に付帯したものと考えることができる．また，人間の感覚を裏切らなければ，値はどのようなものでも良い．そこで便宜上類似性を示す距離を定義し，類似性が失われるに従ってこの距離が増大すると考え，人間の類似性感覚を表１のように数値化した．論理積や論理和を含む意味構造は図１のように全体の指標を求めた．

表１　類似性の感覚と意味的距離指標値

指標値	類似性の感覚	指標値	類似性の感覚
0	（ほとんど）等価	5	半分は似ている
1	ほとんど似ている	7	少しは似ている
2	大部分似ている	10	あまり似てない
3	多くは似ている	20	ほとんど似てない

Generation method of keyword retrieval commands in IRIS

Kohji AKIYAMA
(FUJITSU Limited)

$$m(A \cap B) = m(A) + m(B)$$

$$m(A \cup B) = \max(\, m(A),\ m(B)\,)$$

ただし　A，B　意味構造
　　m(A)　Aの意味的距離指標値

図1　論理積・論理和における意味的距離指標の求め方

(3)意味的距離指標値による戦略動的生成

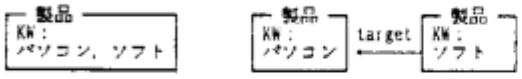
なるべく元の意味構造に類似になるように検索戦術を適用することは，検索要求に合致したものから検索するという目的にかなった妥当なヒューリスティクスであると言える．そこで，戦術適用規則に付帯して記述した意味的距離指標を，戦術の適用順序の動的決定にも利用した．指標値の小さい戦術から適用するような戦略を見つけることで，検索要求に合ったものから検索することができる．

5．最初の評価

現在のＩＲＩＳ拡張版は，質問文解析部およびＡＲＥＸの基本部分の試作が完了した段階である．最初の評価として，対象分野の典型的な検索要求を表す日本語質問文数例と，ＩＲＩＳプロトタイプ版で用いた約八百件のテキストベースとを使って，ＡＲＥＸの小規模な評価を行った．この結果，以下に示す改善が判明した．

(1)戦術適用条件の記述性・精密性向上

分野モデルの拡張によって，特に，下位語や関連語に展開する用語戦術，および，不要語を削除する縮約戦術について，名詞的概念間の関係を用いたより精密で可読性の高い記述が可能になった．例えば，「パソコンソフト」という記述に対して生成される意味構造は，プロトタイプ版では図2(a)，拡張版では図2(b)となる．縮約戦術の適用規則を記述した場合，図2(a)では「製品と製品の並びでは後の単語の削除不可」だが，図2(b)では「製品の対象格が有れば単語の削除不可」となる．このような改善により，「ＩＢＭの最新大型機は何時出るのか」という質問文に対する最終結果件数は，プロトタイプ版では17件，拡張版では6件となり，再現率はどちらも同じだが適合率の向上（ゴミ情報の件数の減少）が認められた．



(a)プロトタイプ版　　(b)拡張版

図2　「パソコンソフト」に対する意味構造

(2)結果件数に依存しない推論停止条件の実現

意味的距離指標値が結果件数に依存しない推論停止条件となり得るかを調べるため，幾つかの質問文に対してＡＲＥＸの検索結果を調査した．この中の2文について，意味的距離指標値を横軸に，再現率および適合率を縦軸にとって表したものを図3に示す．図においてそれぞれ(a)，(b)は「ＣＡＤ分野に進出した半導体企業」および「ＩＢＭの最新大型機は何時出るのか」という質問文に対する検索結果であり，回答となると思われるテキストは八百件中それぞれ22件および3件である．前者は回答件数が比較的多い例であり後者は少ない例である．図3では，(a)(b)のどちらも指標値が11を超えるあたりで最良の結果となっている．このことは，推論停止条件としての意味的距離指標の妥当性を示唆していると思われる．

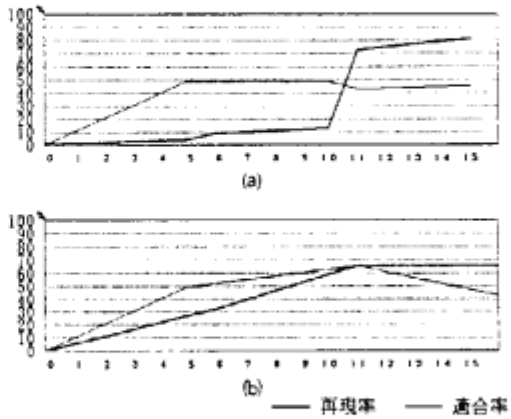


図3　最終指標値に対する再現率および適合率の変化

(3)最適な戦略の生成

プロトタイプ版の固定的戦略では，まず名詞的概念に対するキーワードの関連語等を集める用語戦術を適用した後で，結果件数が少ない場合にはその概念に対するキーワードを抹消する削除戦術を適用するため無駄が生じた．拡張版では，意味的距離指標値を停止条件としたので，削除戦術を使っても目標とする指標値以内になる場合には，削除されるキーワードに対する用語戦術を省いて無駄を排除している．また，ある戦術を適用すると別の戦術を適用できなくなる場合，拡張版では動的戦略生成の一環として，黒板を2つの状態に分岐させてそれぞれ適用するようにしたため，プロトタイプ版に比べてより良い検索式を生成することが可能になった．

6．今後の課題

今回の評価はまだ小規模なものであり，今後，質および量を大規模にして以下に示すような評価・検討を行う必要がある．

①多くの質問文と数千件のテキストベースとを用いた評価．

②各戦術適用規則に対する意味的距離指標値の設定基準の検討．

③結果の予測を用いた計画立案処理による戦略生成方式の検討．

謝辞　本研究は第5世代コンピュータプロジェクトの一環として行われ，ＩＣＯＴ第2研究室の内田，吉岡の両氏を初めとする方々に御支援頂きました．ここに印して感謝いたします．

【参考文献】

[Bates 79]　Bates, M. J. "Information Search Tactics". Journal of the American Society for Information Science, pp. 205-214. 1979.

[杉山 86]　杉山，秋山，伊吹，川崎，内田．「自然言語理解に基づく情報検索システムIRIS」，情報処理学会自然言語研究会資料58-8．1986.

[秋山 87]　秋山，杉山．「従来型情報検索システムへの知的インタフェースとしてのIRIS」，情報処理学会第34回全国大会論文集，pp. 1327-8．1987.

[秋山 88]　秋山．「テキスト情報の知的検索における諸問題」，情報処理学会ＤＢシステム研究会資料64-3．1988.