

新聞記事検索における観点を考慮したクエリー拡張手法

高野 敦子 (兵庫大学)

平井 誠 (大阪市立大学 工学部)

計算機による情報検索が広く普及し、精通していない領域での検索をユーザ自身が行うケースも増えてきている。それに伴い、ユーザが求める漠然とした情報を具体化するという初期の段階での支援を実現するシステムが求められてきている。我々はそのような状況を踏まえて、ユーザの自然言語を用いた自由な入力から有効な検索キーワードを生成する支援システムを提案している。本稿では、有効なキーワードは検索の「観点」に依存することに着目し、ユーザが指定する「観点」に対応する分野の専門事典情報呼び出すことにより動的な語彙間の関連付けを実現する枠組みを示す。これは、「観点」に応じて必要な索引語のみの必要な方向への拡張を可能にする。本稿では、日経新聞データベースの検索に対する実験結果を示すが、語彙間の関連付けは必要な語彙間の関係のみを予め対象データベースとの照合によって各専門事典情報上で活性化しておくもので、機械的な処理で実現するため、一般のデータベースにも柔軟に対応可能な枠組みである。

Query Expansion Based on User's Point of View

Atsuko Takano Economics and Information Science Hyogo University

Makoto Hirai Faculty of Engineering Osaka City University

The purpose of this paper is to propose a framework for making use of several specialized cyclopedias for information retrieval man-machine dialogue system. The system does not require user to formulate his query.

It explores effective keywords from user's point of view based on his input.

In order to realize the scheme, it uses dynamic relations between index words, by using the cyclopedia which is selected according to the point of view.

1 はじめに

一般に情報検索作業において、要求の具体化が最も困難な作業の1つであることは既に報告されている^[oddy]。計算機による情報検索が広く普及し、検索の対象分野や検索作業自体に精通していないユーザも直接計算機を操作して検索を行うケースが増えてきている近年、ますますユーザの要求の具体化に対するシステムの支援が求められるようになってきた。それに対する1つのアプローチとして我々は、ユーザが

自然言語を用いた自由な入力によって検索条件を示すことを可能にする検索支援システムを提案してきた^[高野]。これによってユーザは、「精通していない分野において、要求情報を適切に表現するキーワードを探す」作業が免除され、大幅な負担削減が期待される。

ユーザが自然言語で表現した入力を、計算機が意味レベルまで理解することは、現在の自然言語処理技術では困難である。しかし支援システムとしては、形態素解析プログラム等の言語表層的意味レベルまで理解することは、現在の

自然言語処理技術では困難である。しかし支援システムとしては、形態素解析プログラム等の言語表層的情報を用いた方法でユーザの入力を自動的に索引付けし、索引語の集合に変換することは可能である。ただし、検索対象の分野に精通していないユーザが入力した表現から表層的に抽出できる索引語のみからは有効な検索ができない場合が多い。そこで、ユーザの要求をよりの確に表現する検索キーワードを生成する必要がある。

このような問題に対して、語彙の類似関係や連想関係を整理したシソーラスや概念体系を用いてユーザの入力中の語彙の類似語や関連語に拡張する研究が盛んに行われてきた^[chutze]。

しかし、汎用のシソーラスなどを用いた場合、検索漏れを少なくするという利点はあるものの、不適切な文書が多く検索され、かえって性能が低下する可能性があることが知られている^[徳永]。シソーラスの中で語彙は、上位/下位、全体/部分などの関係によって体系的に関係付けられているが、索引語を拡張する際にたどる関係と方向を決めることが一般には困難であることが、シソーラスを有効利用できない原因と考えられる。

これに対して、拡張する方向が分野に依存するという考えから、分野専用のシソーラスを構築し、それを用いて効果をあげている研究がある^[徳永]。このことをさらに一般化して考えると、たどるべき語彙間の関係は検索の「観点」によって変わってくるということであり、その意味から拡張に使用する語彙間の関係は従来のような静的なものではなく、観点によって動的に変化する必要がある。

観点を定めることによって、見当はずれの拡張は排除できるものの、拡張に使用すべき関係をユーザの入力から予め認定することは困難である。例えば、1999年の日経新聞の記事データベースにおいて、「石綿」をキーワードとすると関連記事を検索できるが、環境分野でその同意語と見なされる「アスベスト」をキーとしてはそれらの記事は

検索できない。ここで、「アスベスト」ではなく「石綿」が使われていることに確たる根拠は見つけれない。ユーザの入力から抽出された索引語が「アスベスト」であった場合、「石綿」に拡張すればいいということは、検索してみて初めてわかることである。つまり、索引語の拡張の問題は、対象とするデータベースの語彙選択に依存し、かつその選択方法に対する明確なルールは一般的には見出しにくいということになる。このことから、シソーラスのように語彙間の関係が体系的に整理されている必要はなく、語彙間に局所的に関係があるという知識のみが有用であると考えられる。

以上のことから、検索の「観点」に応じて拡張すべき語彙間の局所的な関係とその方向が活性化するような、動的な索引語間の関連付けを実現する仕組みが必要となることがわかる。ここで、検索時にユーザが検索の「観点」として対象分野などを指定することは一般的には難しいことではない。しかしながら、「観点」に応じて動的に変化する語彙間の関係を記述した辞書を作成することは多大な労力を必要とする。

この問題に対し我々は、最近電子化が急速に進んでいる各分野の専門事典に注目した。専門事典には多くの場合、その分野で重要な語彙を見出し語とし、その解説および、見出し語と見出し語間の関係が記述されている。「観点」に対応する分野の電子化された専門事典が存在すれば、そこに収められている見出し語を軸とした語彙間の主要な関係は比較的容易に抽出できる。検索時にユーザの指定した「観点」に応じた専門事典を呼び出し、そこから抽出できる語彙間の関係を使うことによって、動的な索引語の関連付けが実現できる。

対象データベースの語彙選択に関するルールが認められない以上、専門事典情報から抽出される関連の中から拡張に使用すべき関係をユーザの入力からシステマティックに認定することは困難である。しかし、対象データベースの語彙選択

に関する個別的な情報を使って検索時に試行錯誤的に語彙間の関連をたどることは避けることができる。今の例を考えると、日外アソシエーツの環境問題情報事典では、「アスベスト」と「石綿」はともに見出し語でそれらの間には同意語としての関連性が記述されている。

そこで、予めこの事典情報の中の「アスベスト」から「石綿」をたどる関係を活性化しておく、具体的には“検索キーワードとして「アスベスト」の代わりに「石綿」を用いよ”という情報を組みこんでおけばよい。このことは、各分野の専門事典情報中のすべての見出し語に対して対象とするデータベースでの検索可能性および、検索可能な語彙との関係を専門事典情報に予め記述しておく処理を、初期設定として1度行っておくことによって実現できる。これによって検索時にはユーザが指定する「観点」に応じた専門事典情報を呼び出し、検索テーマに対する記述力が高いと考えられる語彙のみを分野に適した観点から対象データベースの語彙選択に則して拡張できる。

本稿では、上記の考えに基づき、各分野の専門事典情報を媒介として、ユーザが自然言語を用いて自由に入力した検索条件から具体的な検索キーワードを導き、検索式を生成する枠組みを提案し、実験によってその有効性と課題を検討する。

実験では約20万記事からなる1999年度の日経新聞から環境問題に関する記事を検索する作業を例題として取り上げ、ユーザの入力例から抽出した索引語から本稿で提案する枠組みに基づく手法を用いて検索式を生成する。その際に利用する事典情報は、日外アソシエーツが提供する環境問題情報事典[nichigai]である。この事典は871の見出し語から成る。実験の際に用いたユーザの入力例としては環境に関する冊子の題名を用いた。これらは既に電子化されており、専門的な用語を用いたものから、一般的な表現のみでテーマを表したものでまで様々な表現方法が使われている点で、本実

験の入力例に適している。

2. 専門事典情報を利用した検索条件の生成

2.1 生成方法の概要

支援システムの処理は、事前に専門事典情報を対象データベースに特化する初期化処理と、検索時にユーザが入力した検索条件から有効な検索キーワードを生成して検索し、その結果を表示する処理の2段階からなる。生成された検索キーワードを使って、具体的にどのような検索手法を適用するかは対象データベースに依存する。今回の実験で用いた日経新聞の記事データベースは、索引語を論理演算子で結合した式で検索条件を表すブーリアン検索機能を備えている。

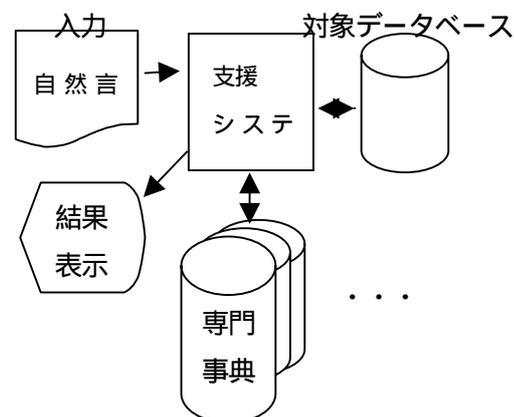


図1. 索引語の拡張方法の概要図

2.2 専門事典情報のデータベースへの特化

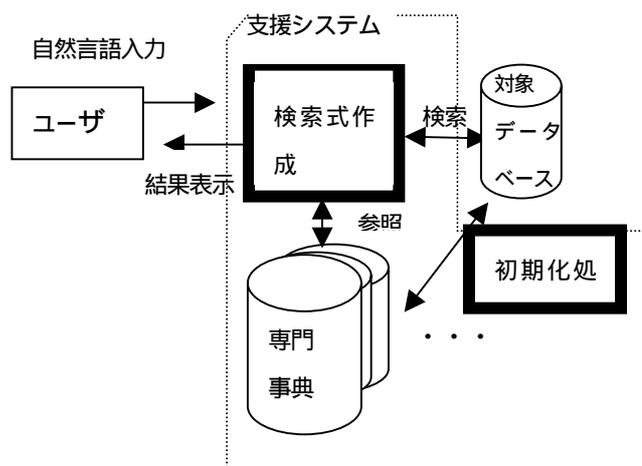


図2. 専門事典情報中の見出し語間に設定する関連づけの概念図

専門事典情報中のすべての見出し語_iに対して対象データベースに特化したキーワード(の集合)

の対応を試みる。これを特化キーワード（集合）と呼ぶことにする。その手順は以下ようになる。

見出し語_iと関連する見出し語_{i1}, 見出し語_{i2}, …を抽出する。

見出し語_iおよび関連する見出し語_{i1}, 見出し語_{i2}, …について対象データベースにおける検索可能性を調べる。

（ア）検索可能な見出し語が存在する場合：
検索可能な見出し語の集合{見出し語_i}を見出し語_iの特化キーワード集合とする。

（イ）検索可能な見出し語が存在しない場合：
見出し語_iに対する特化キーワード（集合）はないとする

各見出し語に関連する見出し語を抽出する手がかりとして、本稿では事典に記述された以下の2つの情報を用いる。

（1）見出し語間にふられた「関連語を示すタグ」

（2）各見出し語に対応する解説文中から表層的処理によって抽出できる同等表現を表す記述

利用できる具体的な手がかりは専門事典毎に多少異なる。しかし、一般的に上記のような語彙間の関係に対する情報は何らかの形で事典に含まれ、表層的な処理で抽出できると考えられる。今回具体的に用いた手法は後述の実験の詳細で述べる。

2.3 特化キーワードによるデータベースの検索

支援システムはまず、ユーザが指定した観点に対応する専門事典情報を呼び出す。次に表層的な処理を用いてユーザの入力から自動的に索引語を抽出し、索引語の集合に変換する。表層的な処理に留めるのは、ユーザが自然言語で表現した単発的な入力を、計算機が意味レベルまで理解することは、現在の自然言語処理技術では困難であることに加えて、情報要求を十分に具体化できていないユーザの入力を深く解析することは意味がないという理由からである。

ここで、抽出した索引語は専門事典情報中の専門性の高い特化キーワードと照合することを考慮して、索引語の基本単位は複合語を許した形式処理によって可能な限り大きな単位とする。今回用いた具体的な抽出方法については、後述の実験の詳細で述べる。

抽出した各索引語を呼び出した専門事典情報の見出し語と照合する。この際、専門性の高い見出し語と基本単位を大きくした索引語の間で照合を図るため、そのままでは特定性は高くなる反面、照合の成功率は低くなる。そこで、照合は部分的な一致も成功とみなす方法を採用する。部分的な一致の具体的な定義は、再現性を重んじるか、精度を重んじるかなど検索の目的によって異なる。今回採用した定義は後の実験の詳細で述べる。

この照合によって、ユーザの入力が特化キーワードの集合に変換される。ここで、入力から抽出された索引語の中で、照合される見出し語がない場合は、検索式には用いられないことになる。これによって例えば「問題」や「視点」などの一般的な広い意味を表し、検索条件を表す上であまり有用でない語彙を排除することができる。

このようにして抽出された特化キーワードを論理和の演算子で結合した検索式によって対象データベースに対してブーリアン検索を行う。ここではユーザの入力に対して意味解析を行っていないので、特化キーワード間に論理的な関係を見出すことができないこと、また、各特化キーワードが専門性が高いので各キーワードが検索する記事数があまり多くないという理由から、全ての特化キーワードを論理和の演算子で結合した。

しかし、例えば「特定物質の規制等によるオゾン層保護に関する法律 その背景、内容及び今後の課題等（環境法の現状と問題点）」という入力において、「オゾン層」と「環境法」の2個の索引語は特化キーワードとなるが、この2個の索引語を比較すると、入力対象とするテーマに関する特

定性が異なる。実際、実験に用いた日経新聞データベースでは、「オゾン層」をキーワードとして検索できる記事が32件あり、そのうち3件は要求情報に一致、残りの29件は関連記事であるのに対して、「環境法」をキーワードとして検索される記事は338件あるが、このうち情報要求に一致した記事が1件あるのみで、あとはほとんど関連性のない記事であった。

このような特定性の異なるキーワードが同等に扱われるのは好ましくない。この問題に対して、本手法では、専門事典情報を対象データベースに特化する初期化処理の過程で各特化キーワードの検索記事数が得られるため、それを利用した対応ができる。この記事数は索引語の特定性を表す尺度として知られているIDF (inverse document frequency) [sparck]における索引語の出現文書数に相当する。このことから、 $df(kw)$ をキーワード kw が検索する記事数とすると、その重み $w(kw)$ を次の式で表すことができる。

$$w(kw) = \log(N/df(kw))+1 \quad \dots (1)$$

この重みを使うことによって、ユーザの入力が特化キーワードの集合 $KW = \{kw_1, \dots, kw_n\}$ に変換されたとき、それを使って検索された各記事 at に対する得点 $P(at)$ を以下の式で与える。

$$P(at) = (KW_{at} \quad kw_i)w(kw_i) \quad \dots (2)$$

ここで、 KW_{at} は at を検索するキーワードからなる KW の部分集合とする。

一般にブーリアン検索の問題点として検索結果に順序付けができないことが知られており、それを改善する拡張ブーリアンモデル[salton]は計算量の大きさが問題になって運用上での適用が難しい。それに対して、上記の方法は非常に簡便的ではあるが、検索時に計算量の負荷をほとんど伴わずに検索結果の順位付けが可能となる。

3. 検索実験の詳細とその結果

3.1 環境問題情報事典を日経新聞記事データベースに特化する実験

環境問題情報事典の主要な内容を表1に示す。

タグ	内容	タグ	内容
ID	ID 番号	WR	“を見よ” 参照語
MD	見出し語	WR	“をも見よ” 参照語
EX	解説	RF	関連語

表1. 環境問題情報事典の主要な内容

各タグに対する内容を形態素解析システム「茶筌」[matsu]を用いて、形態素解析する。見出し語は形態素のリストとして表す。解説部は簡便な方法で索引付けを行い、索引語の集合として表す。簡便な方法とは、形態素に分割した後、連続した名詞の結合など可能な結合を施して名詞句を抽出し、その中で代名詞などの索引語に適さない品詞の語彙を除去したものを索引語とする方法である。この実験では、論点を絞るために索引語を名詞句に限定する。

その結果を用いて、全見出し語871に対して以下の方法で関連する見出し語を抽出した。

(1) 参照語、関連語を示すタグ: WR, MR, RF がふられた見出し語を抽出する。

(2) 解説部の形態素解析結果中で、「とも」(助詞-副助詞)と「いう」(動詞-自立)の連続出現を検出し、それに先行する索引語を抽出する。

この方法で抽出した関連見出し語に対し、前述の方法で検索記事数を調べ特化キーワード(集)を対応させた。特化処理を行う前は、871個の見出し語の中で418個がそれ自体では記事を検索できなかった。その中で、228個に対して関連語を特化キーワードとして対応させることができた。今回の実験では、初期化処理によって、「環境」という観点を設定したとき対象データベースに対して有効な検索キーワードとみなせる基本索引語を453個から681個に拡張できたことになる。

3.2 特化キーワードを用いたデータベースの検索

ユーザの入力例200件に対して検索実験を行った。ユーザの入力は前節で述べた環境問題情報事

典の解説部と同様の方法で索引語を抽出し、索引語の集合に変換する。例えば、「建築物に使用されたアスベストの撤去をどうすすめるか」は次のような索引語の集合に変換される。

{ 建築物, 使用, アスベスト, 撤去 }

これらの索引語を専門事典情報の見出し語と照合する。前述のように、この際の照合は部分一致を許した拡張照合とする。具体的には、見出し語と索引語をそれぞれ形態素列で表し、連続する2個以上の形態素が一致した場合に照合に成功したとみなす。例えば、索引語：“温暖化問題”は(温暖, 化, 問題), 見出し語：“地球温暖化”は(地球, 温暖, 化)とあらわすことができ、(温暖, 化)が一致するため、照合に成功する。

この方法で照合した結果、各索引語は表2のような結果となった。表中の検索数は、予め専門事典情報に記述してある値である。

索引語	照合結果	検索数	特化キーワード
建築物	成功	365	(建築物)
使用	失敗	0	
アスベスト	成功	5	(石綿)
撤去	失敗	0	

表2. 照合結果

これによって検索式は、照合に成功し、かつ検索数が0でない2個の索引語に対する特化キーワードの論理和結合となり、【建築物 OR 石綿】となる。

日経新聞記事データベースは、既に索引付けが行われているので、今回の実験の検索処理では、その索引付けを利用した。ブーリアンモデルを用いた検索には予め用意されている転置ファイルを用いた。記事の抽出プログラムは言語処理学会が提供しているドライバーを用いて作成した。

この例に対して 本手法は次の3つの効果を示した。

(1) 有効な検索キーワードの生成

索引語「アスベスト」から有効な検索キーワード

「石綿」を生成して、検索条件を満たすと考えられる記事の検索数を1件から5件に増やした。

(2) 不要語の処理

入力文から抽出された索引語の内、「撤去」は見出し語と照合されないため、検索式には用いられない。「撤去」は57件の記事を検索するが、その中で環境に関する記事は5件のみであり、そのうち、石綿に関連する記事は0件である。これは「撤去」が「環境」という観点からは、意味が非常に一般的で有効な検索キーワードとは成り得ない語彙であることを示している。このことから、「環境」という観点から検索キーワードとして不適切な不要語と呼べる語彙を排除できていることがわかる。

一方、それ自体記事を1件も検索しない索引語「使用」は、「撤去」と同様の理由で不要語の扱いとなる。このことは、「使用」の検索式からの排除は検索結果に影響は与えないが、前もって検知される不要な検索を排除することになり、検索処理の効率化を助けられていると考えられる。

(3) 検索記事の順位付け

キーワード	件数	適合数	得点
「建築物」&「石綿」	1	1	19.1
「石綿」のみ	4	4	11.8
「建築物」のみ	365	0	7.3

表4. 検索結果に対する順序付け

不要語を取り除いた結果、「建築物」と「石綿」の2つのキーワードの論理和で検索が行われ、検索される記事に対して先に示した(2)式によって表3のように得点が計算される。得点の高い順に表示することによって適合記事が先に表示されるなどの効果的な結果表示が可能になる。例えば今回の例であれば、上位5件に絞って表示すれば、その中にすべての適合記事が含まれることになる。

4 実験の評価および問題点の検討

今回の実験全体の評価および問題点の検討結果を以下の2点について述べる。

4.1 索引語の拡張

入力文 200 件から抽出されたのべ索引語 9890 個のうち、そのままでは 1 件も検索できないものが 5663 個あった。その中で本手法を用いて特化キーワード（集合）を生成し関連記事を検索できたものは、534 個であった。

最初に記事を検索できなかった 5663 個の索引語のうち環境に知識を有する人間の判断で、検索条件を明らかに記述しているとみなせるものは 1133 件あった。これは対応する特化キーワードを生成できた 534 個を全て含んでいた。したがって今回の実験では、検索条件を表現するために有効であるが、対象データベースの語彙選択に適さない語彙の中の 47% を対象データベースに適したキーワードに変換できたことになる。

ここで、この変換率は、使用した専門事典の見出し語の豊富さに依存する。つまり、支援システムが用いる専門事典をさらに見出し語の豊富なものと取りかえることにより、この変換率を上昇させることが可能である。したがって、単独で要求情報を記述できるような高い特定性を持つ索引語については、変換率を上げることは比較的容易である。

これに対して新たな解決策を必要とする索引語は、意味が非常に一般的で、一般的には有効なキーワードとなり得ないが、「環境」という観点を設定すると、有効なキーワードとなり得るものである。例えば、「地球にやさしいライフスタイル」という入力から直接抽出される「ライフスタイル」はそのままでは環境に関する記事を検索できない。ここで、観点を「環境」に特化して「エコライフ」と拡張すると、関連記事が 10 件検索できる。ここで例えば、「ライフスタイル」と「環境」を論理積の演算子で結んだ式で検索しても、この 10 件は検索できない。「ライフスタイル」という一般的な語彙から観点を設定して「エコライフ」のような専門性の高い語彙を生成することは重要な問題と考えられる。この語彙間の関連付けには、見出し語

とその解説部との関係が利用できる。専門性の高い“見出し語”を比較的一般的な表現で説明したのが“解釈部”である。実際、事典情報の中で「ライフスタイル」は見出し語「エコライフ」の解説部の中に次のように出現する。

\\MD\Eコライフ (ecolife)

\\EX\Eコロジカル・ライフスタイルの略称。いわゆる「環境にやさしい暮らし」の意。... 具体的には、同庁は冷暖房の節約・マイカー使用の削減・リサイクル促進などを推奨した。

このような関連付けを抽出する具体的な実現方法の確立は、今後の課題であるが、一般用語から関連する専門用語への拡張の可能性を持つという点からも専門事典情報の有効性が示された。

4.2 不要語の処理

ユーザの入力から抽出される索引語の中で、検索キーワードとして採用しないものを不要語と呼ぶと、これらは以下のように 2 つのカテゴリーに分類される。各カテゴリー毎に生じる不要語削除の効果を以下にまとめる。

(1)【関連のない記事を多く検索する索引語で、これらを排除することにより検索結果の精度が高まる。】このカテゴリーに属する不要語は観点(対象分野)に依存する。例えば入力例から抽出された「地球」は、「環境」という観点からは重要語とはみなされないが、「天文学」などの観点からは、重要な検索キーワードとみなされるべきであろう。今回の実験では、記事を検索できる索引語 4227 個の内、3411 個がこのカテゴリーに属する不要語として削除された。

(2)【対象データベースから 1 件の記事も検索できない索引語で、これらを検索前に削除することにより検索の効率化が図れる。】今回の実験では、9890 個の索引語のうち、その 57% にあたる 5663 個がこのカテゴリーに属する。先に述べたように、次のステップとしては、これらを「環境」という観点に特化した有用な検索キーワードに拡張することを検討する必要がある。

上記のような効果に対して逆に、第1の 카테고リーに属する不要語の削除に対しては、有効なキーワードとなり得る索引語まで削除してしまうという問題点も生じる。この 카테고リーに属する索引語 3411 個は、次の3種類に分類できた。

(あ) 非常に一般的な概念を表す語彙で、関連のない記事を多数索引する(問題、計画、生活など)

(い) 一般的な概念を表す語彙で、関連の無い記事も多く検索するが、対象記事を絞り込む機能は持つ。(首都圏、有機化合物、資源化など)

(う) 具体的な語彙で、有効な検索キーワードになる可能性が高い。年代や地名・人名のような固有名詞など(チェルノブイリ、イタイタイ病など)

(い) の 카테고リーに属する索引語は、再現率を高める可能性を持つが、今回の実験で採用している検索式の構成方法では精度を非常に悪くする可能性が高い。(う) の 카테고リーに属する索引語は、一般的に検索記事数が少ないので、適当な上限値を設定し、検索記事数がそれ以下の索引語は検索キーワードとして採用するという解決方法が考えられる。今回の実験で例えば、上限値を10件に設定すると、733 個が新たに、特化キーワードに加わることになる。この方法は非常に簡便であるが、再現率が確実に上昇することに加えて、検索数の少ない索引語であることから、精度の下があまり問題にならないという長所を持つ。

5. おわりに

各専門分野の事典情報を利用して、ユーザが自然言語を用いて自由に入力した検索条件から有効な検索キーワードによって構成される検索式を生成する枠組みを提案した。

観点毎に、対応する分野の電子化された専門事典を対象データベースに特化した専門事典情報を用意し、ユーザが検索時に指定する観点に応じた専門事典情報を用いて検索支援を行うことにより、観点に応じた検索条件の定式化が行えることを実験によって示した。

検索の観点は、それに適した電子化された専門事典があれば、対象データベースに対して初期設定のみで随時追加できる。初期設定は使用タグなどの簡単な変更のみで機械的に処理できる。さらに、専門事典を取りかえることによって、検索目的に応じた再現率や精度の調整も可能である。

また、対象データベースもブーリアン検索機能が備わっていれば、専門事典情報を容易に取りかえることができる。このように、本稿が提案する枠組みは柔軟で拡張性の高い枠組みと言える。

参考文献

- [Salton 88] Salton, G. Buckley, C.: Term-weighting approaches in automatic text retrieval. *Information Processing Management* 24(5), pp.513-512(1988)
- [Matsumoto 97] 松本裕治, 北内啓, 山下達雄, 平野善隆, 今一修, 今村友明: 日本語形態素解析システム「茶筌」 version 1.0 使用説明書(1997)
- [Tokunaga 99] 徳永健伸: 情報検索と言語処理(第5巻), p.34 東京大学出版会(1999)
- [Oddy 77] R.N.Oddy, : Information Retrieval through Man-Machine Dialogue, *Journal of Documentation*, Vol.33, No.1, .pp.1-14(1977)
- [高野 96] 高野敦子, 平井誠, 北橋忠宏, : ユーザの興味を考慮し発想支援を目指した応答生成手法について, *電子情報通信学会技術報告*, Vol.96 No4.4, .pp21-26(1996)
- [Frakes 92] Frakes, W.B. Baeza-Yates, R. : *Information Retrieval: Data Structures Algorithms*. Prentice Hall(1992)
- [Schutze 97] Schutze, H and Pederson .J .0 : A coocurrence-based thesaurus and two applications to information retrieval, *Information Processing Management* 33(3), pp.307-318(1997)
- [Sparck 72] Sparck Jones, K : A statistical interpretation of term specific cooccurrence-based thesaurus and two applications to information retrieval, *Information Processing Management* 33(3), pp.307-318(1997)
- [Nichigai 97] 環境問題情報事典, 日外アソシエーツ(1992)