

オントロジー高度利用のための 多段階展開型検索手法の提案

A Proposal of Multi-Level-Expansion Search for Advanced Use of Ontology

北河 祐作¹ 古崎 晃司¹ 溝口 理一郎¹

Yusaku Kitagawa¹ Kouji Kozaki¹ Riichiro Mizoguchi¹

¹ 大阪大学産業科学研究所

¹ The Institute of Scientific and Industrial Research (ISIR), Osaka University

Abstract: Recently, various ontologies and applications based-on them are developed. For ontology-based applications, it is very important to acquire appropriate concepts requested by the system developer. In this article, we propose “Multi-Level-Expansion Search” for properly acquisition of concepts from ontologies according to the developer’s intentions.

1 はじめに

高度な意味処理を計算機上で実現するために、オントロジー工学の研究が進められており[1]、様々な領域においてオントロジーの構築や、構築されたオントロジーを用いた知識システムの開発も進められている。

オントロジーベースの知識システム開発において、オントロジーに記述されている内容を適切に利用するためには、システムが必要とする情報をオントロジーから適切に取り出すことが不可欠である。すなわち、オントロジーを対象に行う意味処理の中でも、オントロジー内に定義されている概念から、概念定義(意味)に基づいた検索は重要な要素技術である。オントロジー内で定義されている概念を検索する際には、オントロジーの概念的性質上、様々なアプローチが存在するが、現状の検索手法の多くは単に指定された条件を満たす概念をすべて返すといった最も基本的な手法に限られている。しかし、ユーザ(システム開発者)が概念を検索する意図には、より多くの観点が考えられる。例えば、「骨の疾患」を検索する上でも、「どのような骨にも起こりうる疾患」を知りたい場合や「肋骨や軟骨などの具体的な骨に起こりうる疾患」を知りたい場合があり、「骨の疾患」だけではどのような意図の検索であるか不明確である。更に、検索の対象は特定の領域のオントロジーに構築されたオントロジーだけでなく、どのようなオントロジーに対しても可能でなければならない。

以上より、オントロジーの領域に依らず共通して利用可能であり、オントロジーで定義された概念を知識管理のインデックスとして、目的に応じて用い

るための検索手法が必要であることがわかる。そこで本研究では、ユーザの検索目的に応じて適切な概念を段階的に検索する「多段階展開型検索手法」を提案する。更に、検索結果を適切に提示するための検索結果提示方法について考察する。

2 オントロジー内概念検索

オントロジーを用いた意味処理において、オントロジーで定義された概念構造に基づき適切な概念や関係を取り出すことが、オントロジーや意味処理の種類に依らず共通する要素技術となる。これは、意味処理の対象となる内容を適切な検索条件として表し、オントロジー内からその条件を満たす概念(や関係)を検索することに相当する。本研究では、このような「オントロジー内概念検索」を対象として、高度なオントロジー利用に向けた適切な検索手法について考察する。

本章では、検索対象をオントロジー内で定義されている概念とし、オントロジーの種類に依らず、概念の定義内容に基づいてどのような検索できるかの基本要素について考察する

2.1 基本的な検索要素

オントロジーにおいて概念は他の概念との関係性により定義される。よってオントロジー内概念における意味検索は、単なるラベル一致を条件とした文字列による検索ではなく、概念間の関係を用いた検索に相当する。オントロジーにおいて概念定義に用いられる関係には、概念間の一般-特殊関係を表す is-a 関係と、各概念の定義内容を表すのに用いられるその他の関係(part-of 関係, attribute-of 関係など)

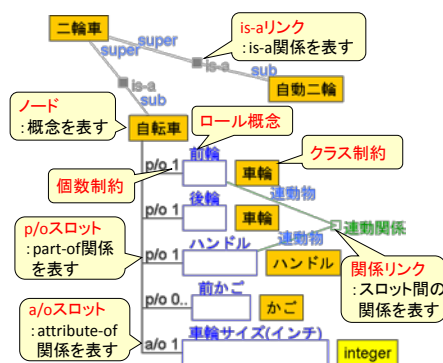


図1 「法造」のオントロジー表現

がある。

図1に「法造¹」を用いたオントロジーの記述例を示す。「法造」が扱うオントロジーは概念を表すノード、概念間の関係を表すリンクおよびスロットを用いて表される。リンクにはis-a 関係を表すis-a リンクがあり、スロットにはpart-of 関係を表すp/o スロットとattribute-of 関係を表すa/o スロットがある。スロットは関係の種類 (p/o またはa/o)、スロットの数を規定する個数制約 (Cardinality)、ルール概念、クラス制約で表される。ルール概念²とは「“もの”が特定のコンテキストのもとで果たす役割を捉えて概念化したもの」で、図の例では車輪が自転車 (コンテキスト) の部品となる時に前輪や後輪の役割を果たすということを表している[2, 3]。ルール概念で定義される役割を担う概念はクラス制約で表され、クラス制約は他の箇所で定義された概念を参照している。よって、オントロジー内の概念検索において検索条件の基本要素として用いられるのは、

- c1) 概念名 (ラベル)
- c2) is-a 階層 (上位/下位概念の指定)
- c3) (スロットで表される) 定義内容

であり、後者2つが意味処理の中心となる。

c2) 「is-a 階層による検索」では、オントロジー上で定義されている概念間の is-a 関係に沿って、指定した概念の上位概念又は下位概念を検索する。is-a 関係には上位概念の定義内容が下位概念に継承され、下位概念では更に詳細な定義内容が追加されるという性質があるため、is-a 階層は検索条件とする概念定義の詳細度の制御に有用である。一方、c3) 「定義内容による検索」では、is-a 以外の関係を用いた概念定義を表すスロットの記述内容を検索条件として指定することで、その条件を満たすスロットを定義

内容に含む概念を検索する。すなわち、c3)は「条件を満たすスロットの検索」と「そのスロットを持つ概念の検索」の2つを組み合わせたものと言える。指定可能なスロットの検索条件を以下に示す。

- s1) ルール概念
- s2) クラス制約
- s3) ルールホルダー
- s4) 個数制約
- s5) スロットの種類 (p/o もしくは a/o)
- s6) スロットの定義内容

これらのうち s1)~s5)はスロットの定義内容の基本要素に相当するのに対し、s6)「スロットの定義内容による検索」は、スロット自身の定義内容も別のスロットを用いて表されるので、c3)「概念の定義内容による検索」と同様に、「条件を満たすスロットの検索」と「そのスロットを持つスロットの検索」を再帰的に組み合わせた検索となる。更に s1)~s3)の検索条件は、それぞれルール概念、クラス制約で参照している概念、ルールホルダーの is-a 階層[4]の情報と組み合わせることで、より柔軟な意味に基づいた検索が可能となる。

以上が、オントロジー内の概念を検索する際の最小単位となる基本的な検索要素である。実際に検索する際は、これらの基本要素を組み合わせることで、より高度な検索が実現される。

2.2 基本的な検索要素の組み合わせ

オントロジー内概念の検索において、意味処理の対象とする概念を取り出す際に必要とされる最も基本的な検索条件は、「指定した概念定義 (意味) を満たす概念の検索」である。これは前節で述べた検索条件のうち、c3)「指定した定義内容による検索」に相当し、定義内容は前節で示した s1)~s6)の条件を用いて指定される。その中でも s1)「ルール概念」と s2)「クラス制約」の組み合わせを指定したスロットの検索が頻繁に行われる。これは OWL では「指定したプロパティの種類と値域 (range) を Restriction として持つ Class」の検索に相当する。

例えば疾患オントロジー[5, 6]において、「骨の疾患」を検索したい場合は、「ルール概念が「異常対象」でクラス制約が「骨」であるスロットを定義内容として持つ概念」を検索すればよい(図2)。

このように概念定義から概念を検索するような場合において、オントロジー内の概念が厳密に定義されている高級なオントロジーを対象にすることで、様々な検索が可能となる。

本研究では、基本的な検索要素の組み合わせのうち、前述の c2)「指定した定義内容による検索」において s1)「ルール概念」と s2)「クラス制約」の組み

¹ <http://www.hozo.jp/>

² 法造では「ルール概念」と「ルールホルダー」という2つの概念を峻別して扱われるが、紙面の都合上、ここではこれらの区別を省略している。

合わせを指定した検索について考察する。

2.3 問題点とアプローチ

オントロジーを用いた意味処理においては、オントロジー内の概念検索の結果が、ユーザ（システム開発者）の検索意図や目的を適切に反映していることが重要となる。例えば、図2の「骨の疾患」という条件で検索を行う際の考えられる意図としては、

- 1) 「骨の疾患」に共通する性質（概念定義）を知るため、検索結果の中でも最も一般的な疾患の検索を目的とし、検索結果として「骨疾患」のみを求める場合
- 2) (後頭骨や脊椎骨や肋骨のような特定の種類の骨にのみ起こる疾患ではなく) すべての種類の骨に共通する疾患の検索を目的とし、検索結果として異常対象のクラス制約が骨（特定の種類の骨は除く）である「骨疾患」、「関節リウマチ」、「骨肉腫」を求める場合
- 3) 後頭骨や脊椎骨や肋骨のようなあらゆる種類の骨を分類した部位における疾患の検索を目的とし、図2で示したすべて疾患を検索結果として求める場合

などがある。“ロール概念が「異常対象」でクラス制約が「骨」であるスロットを持つ概念”という「検索条件を満たす」という点だけを考えると図2で示した全ての検索結果を返すべきである。しかし、1)~3)のようなユーザの意図を考慮すると、検索結果として返すべき範囲が異なり、単に上述の条件を満たすというだけでは、ユーザの意図が不明確である。また検索結果の中にユーザが求める情報が含まれていたとしても、膨大な検索結果が返ってきた時などには、求める結果の有無がユーザにとって不鮮明である場合も考えられる。

そこで本研究では、ユーザの意図に応じた検索に対応するために、概念の is-a 階層をユーザの意図に応じて段階的に展開することによる多段階展開型検索手法を提案する。更に、得られた検索結果がどのような過程で得られた概念であるかを利用することで、適切に検索結果の提示を行う検索結果提示方法の考察を行う。

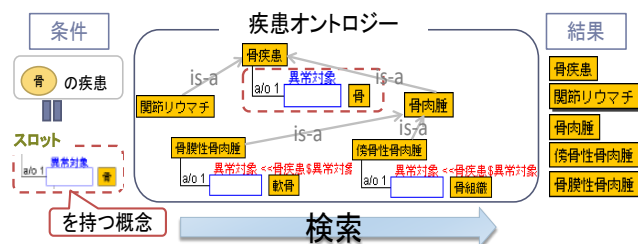


図2 「骨の疾患」の検索例

3 多段階展開型検索手法

オントロジー内概念検索において「指定した定義内容による検索」を行う際には、検索条件としてスロットで表わされる概念定義を指定する。この検索条件と、その条件から得られる結果のそれぞれに対して is-a 階層に沿った下位概念へ段階的に展開を行うことで、ユーザの意図に応じた検索を実現する。更に、それぞれの展開を組み合わせることで、検索手法を複数の段階に細分化し、それぞれを検索目的の範囲に応じて区分する。

3.1 結果展開

検索条件を満たす結果として得られた概念の下位概念には、is-a 関係に沿って定義内容が継承されるため、これらすべての下位概念もオントロジー的に検索条件を満たす。このように、検索結果として得られた概念の is-a 階層に沿って下位概念を展開し、それらの下位概念も検索結果とすることで、結果として返す概念の範囲を操作することを結果展開と呼ぶ。例えば、様々なスポーツに関して定義されている「スポーツオントロジー」において「ボールを用いるスポーツ」という条件で検索した場合、結果展開を適用することで、「球技」の下位概念である「野球」や「サッカー」なども結果とする。

更に、スロットの継承・特殊化のされ方を考慮することで、結果展開を3レベルに区分することが可能となる。スロットの特殊化とは、下位概念において上位概念から継承された定義内容をより詳細なものに書き換えることである。図3の「大動脈疾患」の例では、ロール概念が「異常対象」でクラス制約が「大動脈」であるスロットが定義されている。その下位概念である「大動脈炎症候群」と「大動脈瘤」にはそのスロットがそのまま継承されており、更に下位概念である「腸骨動脈瘤」、「腹部大動脈瘤」、「胸部大動脈瘤」では、継承されたスロットの「異常対象」のクラス制約が「大動脈」から「腸骨動脈」や「胸部大動脈」に特殊化されている。なお、スロットが特殊化されていない場合も、上位概念から継承された定義内容は満たしているため、検索結果としての条件は満たしている。

このスロットの継承・特殊化を考慮した結果展開

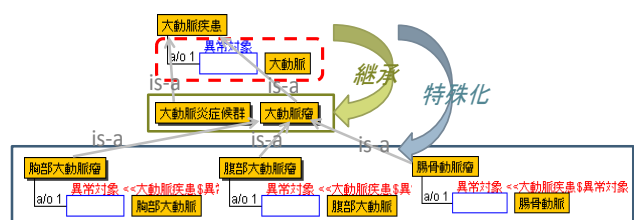


図3 スロットの継承・特殊化

のレベル分けを以下に示す。

レベル1

- 結果展開を用いない
- 検索条件を満たす最上位概念のみ取得

レベル2

- 検索条件となるスロットをそのまま継承している概念まで展開
- 該当スロットが特殊化されている概念は展開しない

レベル3

- 検索条件となるスロットが特殊化されている概念も展開
- レベル1で得られる概念の下位概念全てを展開することに相当する

図3のオントロジーにおいて「大動脈の疾患（「異常対象」のクラス制約が「大動脈」である疾患）」という条件で検索した際には、結果展開レベル1では、「異常対象」のクラス制約が「大動脈」であるスロットが定義されている「大動脈疾患」のみを結果とする。結果展開レベル2を適用すると、検索条件と一致するスロットを継承している「大動脈炎症候群」と「大動脈瘤」も結果とし、更に結果展開レベル3を適用すると、そのスロットが特殊化されている「腸骨動脈瘤」、「腹部大動脈瘤」、「胸部大動脈瘤」も結果とする。レベル1では、条件を満たす概念の中で最も一般的な概念のみを検索し、レベル2では、条件を満たす中でも共通な性質で分類される概念を検索し、レベル3では、条件を満たす概念全ての検索を行う。

3.2 条件展開

検索条件として「クラス制約で参照している概念」を指定したとき、「その下位概念をクラス制約で参照している」場合も、オントロジー的に条件を満たしている。よって、「指定したロール概念とクラス制約の組みを定義内容にもつ概念」を検索する際には、検索条件で指定したクラス制約で参照している概念のis-a階層を用いて、その下位概念をクラス制約で参照しているロール概念を定義内容に持つ概念も検索結果に含まれる。このように検索条件で参照している概念のis-a階層³を展開し、検索条件を満たす概念を検索する手法を条件展開と呼ぶ。例えば、「ボールを用いるスポーツ」という条件で検索したとき、ロール概念「使用道具」のクラス制約が「ボール」であるスロットを定義内容に持つ概念を検索する。そ

³ 本稿では、クラス制約で参照している概念のみを対象として議論しているが、ロール概念やロールホルダーのis-a階層でも同様な議論が可能と思われる。

の際に、条件展開を用いずに検索すると、「使用道具」のクラス制約が「ボール」と定義されている概念である「球技」のみを検索結果とする。一方、条件展開を適用すると、クラス制約で参照している「ボール」のis-a階層を用いることで、「ボール」の下位概念である「野球ボール」や「サッカーボール」を「使用道具」のクラス制約で持つ「野球」や「サッカー」なども検索結果として得られる。このように、条件展開を用いることで、一般的な条件での検索と詳細な条件での検索に段階付けることが可能となる。

なお、条件展開では、検索結果となる概念のis-a階層、すなわち、上位概念から継承された定義内容は考慮しない。よって、上位から継承されたスロットが検索条件を満たしていたとしても検索結果には含めない。

3.3 条件展開と結果展開の組み合わせ

3.3.1 結果展開/条件展開で得られる検索結果の違い

検索結果の違い

結果展開および条件展開を適用した際に得られる検索結果の違いについて、それぞれの展開で利用するis-a階層構造に注目して考察する。

ここで、元の検索条件で指定された概念をクラス制約で参照しているスロットを定義内容に持つ概念をルートとしたis-a階層を主条件参照階層、条件展開により展開された概念（元の検索条件で指定された概念の下位概念）をルートとしたis-a階層を下位条件参照階層と呼ぶ。結果展開（レベル3）を適用した際に得られる検索結果の範囲は主条件参照階層であるのに対し、条件展開を適用した際に得られる検索結果の範囲は主条件参照階層および下位条件参照階層のルートとなる概念となる。よって、すべての下位条件参照階層が主条件参照階層の部分階層として含まれるときには、条件展開で得られる結果はすべて結果展開で得られる結果に含まれる。一方、条件参照階層に含まれない下位条件参照階層が存在するときには条件展開による結果展開では得られない検索結果が得られる(図4)。

すなわち、結果展開で利用するis-a階層が条件展開で利用するis-a階層と異なる分類視点で構築され

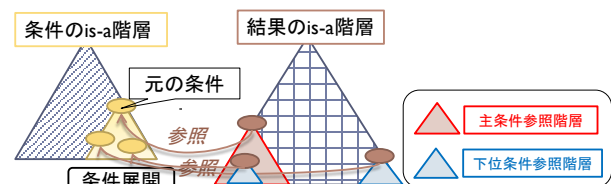


図4 条件と結果のis-a階層構造が一致しないときの条件展開によって得られる結果

ており、条件展開により結果展開で得られない検索結果が得られる。

3.3.2 組み合わせによる展開パターンと得られる検索結果の範囲

本項では、条件展開と結果展開の組み合わせによる展開パターンについて考察を行う。条件展開は適用する/しない場合の2種類、結果展開はレベル1~3の3種類存在するので、展開パターンの組み合わせは表1に示す6種類が考えられる。それぞれの展開パターン適用した際に検索結果として得られる範囲については図5のようになる。

まず、条件展開を用いない(1)~(3)の展開パターンは、3.2節で述べたレベル1~3の結果展開と同様であるので、(4)~(6)の展開パターンについてのみ以下に示す。

(4)「条件展開：○，結果展開：×」

(4)では、条件展開を適用し、主条件参照階層および下位条件参照階層のルートとなる概念も検索結果として得る。結果展開は、展開を適用せずに条件を満たす概念の最上位概念のみを検索結果とするレベル1であるので、条件展開によって得られた各階層のルートである概念のみが検索結果となる。

(5)「条件展開：○，結果展開：△」

(5)では、(4)の条件展開の適用で得られた主条件参照階層および下位条件参照階層のルートとなる概念に加え、それらの検索結果に結果展開をレベル2(条件となるスロットをそのまま継承している概念まで展開する)で適用する。よって、(5)では範囲外であった検索条件となるスロットをそのまま継承している下位概念も検索結果となる。

(6)「条件展開：○，結果展開：○」

(6)では、(4)の条件展開の適用で得られた主条件参照階層および下位条件参照階層のルートである概念に加え、それらの検索結果に結果展開をレベル3(下位概念全てを展開する)で適用する。よって、条件展開によって得られた各階層のルートである概念の

下位概念を全て検索結果とする。

これら6種類の展開パターンで得られる検索結果の範囲を比較すると、(5)と(6)は常に同一となり、更にすべての下位条件参照階層が主条件参照階層の部分階層として含まれる場合においては、(3),(5),(6)で得られる検索結果の範囲も同一となる。

3.4 展開パターンの段階区分の考察

次に6種類の展開パターンを検索目的の範囲に応じた段階区分の考察を行う。まず、(4)「条件展開：○，結果展開：×」の検索は、通常は不必要であると考えられる。なぜなら、(4)の検索では、上位からそのまま継承され定義内容が検索条件を満たしている概念は省かれるためである。ただし、(4)の検索結果となる下位条件参照階層のルートとなる概念の存在は、オントロジー内の is-a 階層上で定義が特殊化される境界となる概念に相当し、有用な情報を持っていると考えられるので、検索結果を提示する際には明示化する必要があると考えられる。

次に、同一の結果となる検索範囲を持つ展開パターンである(5)と(6)を同一の段階に区分する。更に、(3)も(5)-(6)と同一の段階に区分できると考えられる。(3),(5),(6)で得られる検索結果が同一となる条件を満たす時は、同一の段階に区分することは自明であるが、それ以外の場合でも(3)は条件を満たす全ての結果を得たい時に用いられると考えられるので、(3)も(5)-(6)と同一の段階に区分できる。

以上より、実際の検索する際には、(1)，(2)，(3)-(5)-(6)の3段階の展開パターンに分けられる。それぞれの検索目的として、(1)は、条件を満たす概念の内、最も一般的な概念の検索を目的とする際に用いられる。例えば、「骨の疾患」という条件で検索する時、全ての骨で起こりうる疾患の中でも最も一般的な疾患の検索を目的とする。(2)は、「骨の疾患」の中でも(後頭骨や脊椎骨や肋骨のような特定の種類の骨にのみ起こる疾患ではなく)すべての種類の骨に共通するといった、共通な性質を持つ概念の検索を目的とする際に用いられる。(3)-(5)-(6)は、その検索条件を満たすすべての概念の検索を目的とする。

表1 展開パターンの組み合わせ

条件展開	結果展開	
×	×	…(1)
	△	…(2)
	○	…(3)
○	×	…(4)
	△	…(5)
	○	…(6)

×: 条件展開無
○: 条件展開有
×: 結果展開レベル1
△: 結果展開レベル2
○: 結果展開レベル3

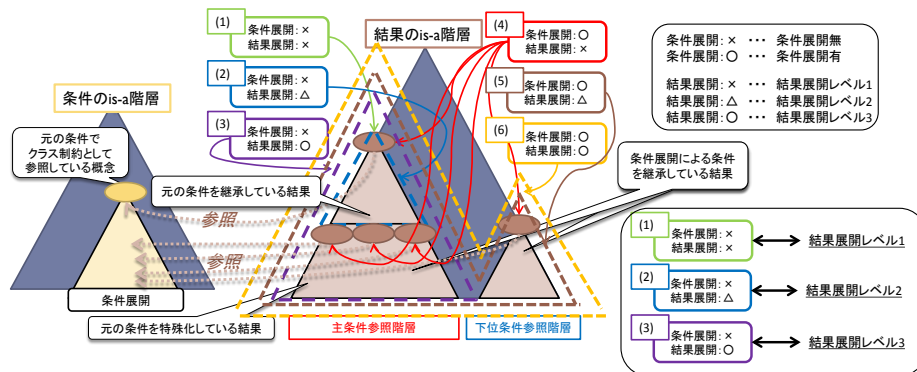


図5 展開パターン毎の検索結果の適用範囲

4 検索結果提示方法の考察

オントロジーに基づくシステム開発を考える際には、オントロジー内概念検索の結果が、システム開発者に意図に即したものであるかが明確となるような検索結果の提示方法が重要となる。それには、得られた概念を単に提示するのではなく、どのような条件で検索された概念であるかを明示することが有用であると考えられる。(オントロジー内概念検索における)条件とは、概念の定義内容であり、概念の意味的な違いは概念の定義内容によって表される。検索結果を提示する際に、概念の意味的な違いを明示することで、システム開発者がオントロジーをどのように利用するかを明白にできると考えられる。

4.1 基本的な検索結果提示方法

検索結果を一覧表示する方法として、is-a 階層を利用したツリー表示が考えられる。また、大量の結果を効率的に閲覧するための方法にソートが考えられる。ソート方法は、ラベル順や結果又はクラス制約で参照している概念の is-a 階層のパスを用いる方法がある。クラス制約で参照している概念の is-a 階層のパスでソートすることで、例えば、疾患の異常対象ごとのソートが可能となる。

4.2 展開パターンを利用した検索結果

提示方法

3章で提案した多段階展開型検索手法の展開パターンを明示化することで、システム開発者にとって意味のある情報の提示が可能と考えられる。ここでは、3.3.2項で述べた展開パターン(4)について考察する。

4.2.1 ルート概念の明示化

3.4節で述べたように、展開パターン(4)の検索で得られる主条件参照階層および下位条件参照階層のルートとなる概念はオントロジー内概念検索において、有用な情報であると考えられる。なぜなら、これらの概念は、条件が詳細化される境界点となっているからである。例えば、「骨の疾患」という条件で検索した際には、条件が詳細化する境界として、「骨の疾患」から「軟骨の疾患」や「硬骨の疾患」、更に「硬骨の疾患」から「肩胛骨の疾患」や「肋骨の疾患」などがある。条件が詳細化されている境界点を明示化することで、ユーザが与えた条件の詳細化が明確になり、どのように条件を設定し再検索するかの支援にも繋がると考えられる。

4.2.2 意味的なツリー展開の考察

検索結果の一覧をツリー表示した際に、概念間の関係や検索条件を考慮した意味的なツリー展開を考える。従来の「法造」のツリー展開は、1階層ずつ又は最下層までの展開であった。それらの操作では、結果一覧の階層が深い時など、求める情報を見つけることが困難である。そこで、展開パターン(4)の検索で得られる情報を用いて、意味を考慮したツリー展開を行う。前項で述べたように、展開パターン(4)によって条件が詳細化されている境界の概念が得られる。この情報を用いて、1度の展開で、この境界の概念まで展開することで、概念の意味的な位置を鮮明にするツリー展開を提示することが可能となる。

5 まとめ

本研究では、ユーザの検索目的に応じて適切な概念を段階的に検索する「多段階展開型検索手法」を提案と、検索結果を適切に提示するための検索結果提示方法について考察を行った。

今後の課題としては、IS-A 関係や p-is-a 関係などのその他の要素にも着目した考察、提案手法を基にした検索システムの実装、それらの評価などが挙げられる。

謝辞

本研究の一部は科学研究費補助金若手研究(A)20680009の助成による。

参考文献

- [1] 溝口理一郎：オントロジー工学，オーム社(2005)
- [2] 古崎晃司，來村徳信，池田満，溝口理一郎：「ロール」および「関係」に関する基礎的考察に基づくオントロジー記述環境の開発，人工知能学会論文誌，Vol.17(3),pp.196-208(2002)
- [3] R.Mizoguchi, E.sunagawa, K.Kozaki and Y.Kitamura: A Model of Roles within an Ontology Development Tool: Hozo, J. of Applied Ontology, Vol.2, No.2, pp.159-179. Sep(2007)
- [4] 砂川英一，古崎晃司，來村徳信，溝口理一郎：コンテキスト依存性に基づくロール概念組織化の枠組み，人工知能学会論文誌，Vol.20 No.6 pp.461-472(2005)
- [5] 国府裕子，周俊，古崎晃司，今井健，大江和彦，溝口理一郎：臨床医療オントロジーに関する基礎的な考察，人工知能学会 第22回全国大会，2E3-01(2008)
- [6] R.Mizoguchi, H.Kou, J.Zhou, K.Kozaki, K.Imai and K.Ohe: An Advanced Clinical Ontology, In Proc. of ICBO, pp.119-122, June 24-26(2009)