

シソーラス「分類語彙表」の検索メニューへの適用

千田 恭子, 篠原靖志, 坂内広蔵

(財) 電力中央研究所 情報研究所

senda, sinohara, bannai@denken.or.jp

1 はじめに

本論では、汎用的なシソーラス ([4] など) を利用した、検索用メニューの構築について検討する。

規程文書や初心者用マニュアルなどの文書は、専門的な内容を平易な言葉で説明しており、その内容や分野に詳しくないユーザむけである。内容に不慣れなユーザがそれらの文書を検索する場合、メニュー選択による方式が便利であろう。

従来の代表的なメニューの自動構築手法としては以下があげられる。

1. 文書名や、文書中の重要語などを選択情報として提示する ([1],[2])
2. 対象文書の内容を、その分野の概念体系に沿って分類する ([6])

しかし、文書内容に予備知識がないユーザにとって、タイトル、重要語、重要概念等の項目名を見ても、その下位項目は想定しづらく、メニュー選択は難しいと思われる。

汎用的なシソーラスの以下の利点を利用すれば、文書内容に不慣れなユーザでも下位項目が想定しやすいメニューが構築できる。

利点 1 特定の内容に特化しない、一般的な意味概念の汎用的な分類であるため、対象文書に対する予備知識がなくても、項目名の意味や体系を想定しやすい。

利点 2 文書から抽出した重要語をシソーラスの分類語に対応づけ、分類項目をメニュー項目として提示することで、簡単に自動構築ができる。

但し、汎用的なシソーラスを検索用のメニューとして使いやすくするには、以下の課題があげられる。

課題 1 項目名として提示された際、下位に分類されている語が想定しにくい項目と、比較的想定しやすい項目がある。[4] のシソーラスを例にとると、想定しにくい項目名として「事柄」「関係」「有無」「様相」、想定しやすい項目名として「親族」「物品」「衣服」「食料」などがあげられる。下位が想定しにくいと、ユーザは選択誤りをしやすい。

課題 2 項目名の表す意味が一般的すぎるため、対象文書での意味が特定できず、選択をためらったり誤ったりする項目名がある。例えば「相手・仲間」という項目名は、契約の「相手」、仕事の「仲間」など多様に意味を想定でき、どれか具体的な意味が想定できないと選択しにくい、もしくは選択誤りしてしまう。

課題 3 シソーラスの階層や項目の数がメニュー提示に適さない場合がある。これに対し [3] では、不要な中間項目を削除して、階層や項目数を調整する手法を提案している。

本論では、ある企業の規程文書に汎用的なシソーラスを適用して、検索用のメニューを構築することが目的である。以降の節で、課題の 1、2 について検討した内容を順に述べる。

2 具象名詞の項目の優先提示

課題 1 について考察するため、「下位に分類されている語が想定しやすい項目」「想定しにくい項目」をとりあげ、その性質の違いを以下に分析した。

想定しやすい 具体物や物理的に計測できるものを指す語の分類項目。

想定しにくい 実体のない抽象的な概念を指す語の分類項目。

これらの項目間で想定しやすさが違うのは何故だろうか。

シソーラスは、意味的に近い語を集めたグループに、それらの語の意味を総括するようなグループ名がつけられているという一面をもつ。下位分類の想定しやすさに差異があるのは、適切な語のグループ分けや名前付けの難しさに差異があると思われる。その視点でみると、上記の「想定しやすい」項目では、語を具体物の通念的な分類や物理的な分類に沿って分類でき、使われる状況に左右されにくい、普遍的なわかりやすい分類を作成しやすいと考えられる。反対に「想定しにくい」項目の語は、指すも

のが実体としては存在しないため、語の意味の直感的な近さだけで分類しやすい。語の意味は、使われる状況によって微妙に変化するため、意味的な近さだけで汎用的に分類するのは難しい。つまり、「想定しやすい」項目は汎用的に分類しやすい語の項目であり、「想定しにくい」項目は汎用的に分類するのが難しい項目であるといえる。

この分析に基づき、具体物や物理的に計測できるものを指す語の分類項目をまず第一階層で優先的に提示し、それ以外の項目は、第一階層で適当な項目が見つからない場合にだけ選択できるように第二階層以下で提示する手法を提案した [5]。これを優先提示メニューと名付ける。

3 目次構造との組合せ

次に課題 2 の解決のために、汎用的なシソーラスの性質を考察した。シソーラスは、特定の内容や分野に特化していない分類であるため、対象文書の内容や分野に予備知識がないユーザにもわかりやすい。その一方で、対象文書とは全く無関係の体系であるため、一般的な意味概念を表す項目名を提示されただけでは、どのように対象文書へたどれるのか想定できず、選択にためらったり迷うことがある。つまり既成の汎用的なシソーラスは対象文書に関する情報が欠けていることが、課題 2 の原因であると考えられる。

この解決には、シソーラスにはない対象文書の(文脈)情報の導入が必要である。そこで本論では、目次の章、節、条などのタイトルや階層は、一種の対象文書の(文脈)情報であると考え、これを汎用シソーラスと組み合わせることを提案する。1節で述べたように目次の項目名等は、文書内容に不慣れなユーザには解りにくい可能性があるが、この問題もシソーラスのように一般的な意味概念の分類体系と組み合わせることで補えると思われる。つまり、汎用的なシソーラスと目次構造を組み合わせた、相補的なメニューの提案である。これを相補メニューと名付ける。

具体的には例えば、シソーラスのメニューから項目を選択した場合、その下位項目が提示されるだけでなく、選択項目に分類されている語が、目次の各文書中の語と幾つ対応するか対応数が表示される。これによって、選択した項目が主にどの目次の文書

の語を分類しているかわかり、分類語の対象文書での意味がある程度想定できる。

同様に、目次メニューから選択した場合も、目次の下位項目だけでなく、選択項目の文書中に出現する語がシソーラスの各項目に幾つ対応するか対応数が表示される。これによって、どのシソーラスの項目と関連が強いかわかり、選択項目の文書がどのような内容について記述しているか想定しやすくする。

4 システムの概要と動作

2、3節で述べた解決手法を組み込んだメニューシステムを、ある企業の規程文書の検索用インターフェースとして実装した。シソーラスの体系には、文献 [4] の国立国語研究所「分類語彙表」を、2節で述べた優先提示メニューの形で利用している。

図 1 がシステムの表示画面例である。画面の左上がシソーラス [4] の優先提示メニューで、左下がその下位項目のメニュー、右上が文書の目次構造を利用したメニューで、右下がその下位項目のメニューである。この画面では、シソーラスのメニューとして、2節で下位分類が想定しやすいと説明した、具象名詞の分類項目を並べた第一階層を表示している。

画面上の上位のメニューから項目を選択した場合、その下のメニューに下位項目が表示される。またその下位項目と、もう一方の体系のより下位のメニュー項目の右側に、3節で説明した、選択項目中の語に各項目がそれぞれ幾つ対応するかという対応数が表示される。複数回項目を選択してきた場合は、全ての既選択項目に対する対応数が表示される。画面下の下位のメニューから項目を選択した場合は、その下位項目全体が上位メニューの位置に移り、元の下位メニューの位置には、新しい下位項目が提示される。それ以外の点は上位のメニューから選択した場合と同じである。このシステムでは、二つのメニュー体系を並行して使用することも、一つのメニュー体系だけを使用することもできる。

4.1 シソーラスからの選択例

「壊れた器材の処分を届け出て買い換えを申請するために、手続きと担当部署について調べる」という検索課題を例に、システムの一連の動作について説明する。

| | |
|------------------|------------|
| 10 シソーラスメニュー | 10 目次メニュー |
| 14 物品 | 1 経営 1 |
| 15 資材 | 2 職務分担 0 |
| 16 衣服 | 3 総務 15 |
| 17 食料 | 4 人事 11 |
| 18 住居 | 5 財務 0 |
| 19 道具 | 6 契約 0 |
| 20 機械 | 7 研究開発 0 |
| 21 性質・材質 | |
| 22 自然・物体・物質 | |
| 10 シソーラスの下位項目 | 10 目次の下位項目 |
| 15.1 資材 6 | |
| 15.2 紙 9 | |
| 15.3 燃料・肥料 11 | |
| 15.4 輪・車・棒・管など 1 | |

図 1: シソーラスの項目「資材」の選択

図 1 は、シソーラスのメニューから、検索課題に関連がありそうな項目「15 資材」を選択したところである。(選択した項目には下線がひいてある。) 3 節で述べたように、この結果より、シソーラスの項目「資材」の分類語が、目次項目の中で最も検索課題に関連が深そうな「3 総務」に多く分布していると解り、項目「資材」の選択は適切であると判断できる。

4.2 目次からの選択例

目次メニューから検索課題に関連が深そうな「3 総務」を選択した結果が図 2 である。項目「15 資材」「3 総務」を続けて選択したため、下位項目には二つの既選択項目に対する対応数が表示される。この情報より、選択してきた項目に対応する語が、シ

| | |
|------------------|----------------|
| 10 シソーラスメニュー | 10 目次メニュー |
| 14 物品 | 1 寄附行為 |
| 15 資材 | 2 権利・職務権限 |
| 16 衣服 | 3 総務 |
| 17 食料 | 4 人事・労務 |
| 18 住居 | 5 財務 |
| 19 道具 | 6 計算機 |
| 20 機械 | |
| 21 性質・材質 | |
| 22 自然・物体・物質 | |
| 10 シソーラスの下位項目 | 10 目次の下位項目 |
| 15.1 資材 6 | 3.1 文書管理規程 8 |
| 15.2 紙 9 | 3.2 技術情報管理規程 0 |
| 15.3 燃料・肥料 11 | 3.3 固定資産管理規程 6 |
| 15.4 輪・車・棒・管など 0 | 3.4 特許取得要領 1 |
| | 3.5 建物管理規程 0 |
| | 3.6 安全管理規程 0 |

図 2: 目次の項目「総務」の選択

ソーラスの項目では「15.1 資材」「15.2 紙」に分類されており、同じ語が目次の項目では「3.1 文書管理規程」「3.3 固定資産管理規程」等に主に分類されていることがわかる。そして、下位項目の「資材」や「固定資産管理規程」は検索課題に関連が深そうであるため、それまでの項目選択が適切だったと判断できる。例えば仮に、目次下位項目の「3.4 特許取得容量」やシソーラス下位項目の「15.3 燃料・肥料」の数値だけが高い場合は、検索課題にあまり関係のない項目を選択してしまったと判断できることになる。

4.3 目次の下位項目からの選択例

目次の下位項目から、検索課題に関連が深そうな「固定資産管理規程」を更に選択した結果が図 3 である。この結果から、今までの既選択項目に該当する語が、シソーラスでは「15.1 資材」だけに分類され、目次の下位項目では最も関連が深そうな「第 3 章 設備、備品」に、主に分類されていることがわかる。ここで、目次の項目「第 3 章 設備、備品」を選択すると、文書の該当箇所が提示され、検索課題に関連する記述を直接探することができる。

4.4 シソーラスの下位項目からの選択例

目次項目を選択せずに、更にシソーラスの項目を選択して、該当する記述を絞りこむこともできる。例えば、シソーラスの項目「15.1 資材」を選択すると、画面は図 4 になる。この図より、検索課題に関連のありそうな「資材担当課」という語が文中にあることがわかる。(シソーラスメニューの末端の階

| | |
|------------------|-------------------|
| 10 シソーラスメニュー | 10 目次メニュー |
| 14 物品 | 3.1 文書管理規程 |
| 15 資材 | 3.2 技術情報管理規程 |
| 16 衣服 | 3.3 固定資産管理規程 |
| 17 食料 | 3.4 特許取得要領 |
| 18 住居 | 3.5 建物管理規程 |
| 19 道具 | 3.6 安全管理規程 |
| 20 機械 | |
| 21 性質・材質 | |
| 22 自然・物体・物質 | |
| 10 シソーラスの下位項目 | 10 目次の下位項目 |
| 15.1 資材 6 | 3.3 第 1 章 総則 0 |
| 15.2 紙 9 | 3.3 第 2 章 土地、建物 2 |
| 15.3 燃料・肥料 11 | 3.3 第 3 章 設備、備品 4 |
| 15.4 輪・車・棒・管など 0 | |

図 3: 目次の項目「固定資産管理規程」の選択

| | |
|----------------|--|
| 15.1 資材 | |
| 15.2 紙 | |
| 15.3 燃料・肥料 | |
| 15.4 輸・重・棒・管など | |

| | |
|------------|---|
| 15.1 使用材料: | 1 |
| 15.1 材料置場 | 1 |
| 15.1 資材 | 3 |
| 15.1 資材担当課 | 1 |

| | |
|--------------|--|
| 3.1 文書管理規程 | |
| 3.2 技術情報管理規程 | |
| 3.3 固定資産管理規程 | |
| 3.4 特許取得要領 | |
| 3.5 建物管理規程 | |
| 3.6 安全管理規程 | |

| | |
|---------------|---|
| 3.3 第1章 総則 | 0 |
| 3.3 第2章 土地、建物 | 2 |
| 3.3 第3章 設備、備品 | 4 |

図 4: シソーラスの項目「資材」の選択

層には、シソーラスの分類語を含む文書中の複合語が割り付けられてある。)この語を選択すると、それを含む文章が表示される。

5 優先提示の有効性の検証

2節で述べた優先提示メニューの有効性を確認する実験を行なった。比較対象として、同じ汎用的なシソーラス [4] に手を加えず直接適用したメニューシステム (仮に直接提示メニューと名付ける) を用意した。両方のシステムで、職員のアンケートなどから集めた検索例題九題を被験者一人に検索させた。具象名詞の分類項目の優先提示で、選択誤りや項目選択回数が減少するか確認する目的で、各システムでの (上位階層への後戻りも含めた) 項目選択回数を数え、どちらが少ない回数で求める情報に到達できるか比較評価した。

結果は、直接提示メニューでの平均 14 回が、優先提示メニューでは平均 10 回に減少した。(但し、検索課題に関連する項目がなくても、最初に提示された優先提示階層から項目を選択しようとして、時間内に終了しなかった例題二つを除く。) よって、ユーザが早期に正しい階層を選択出来れば、優先提示メニューは、項目選択の回数を減少し得ると考えられる。

6 まとめ

本論では、規程文書など平易な用語で専門的内容を説明する文書を対象に、対象文書の内容に予備知識がないユーザでも検索しやすいメニューを、汎用

的なシソーラス [4] を適用して自動構築する手法について述べた。適用課題として、(1) 下位に分類されている語が想定しにくい項目、(2) 対象文書での意味が特定しにくい項目の二つをあげ、1) 通念性が高く下位語が想定し易い、具象名詞の項目を優先的に提示する、2) 対象文書の情報を補うため目次構造をもとにしたメニューと組み合わせる、という手法を提案し実装した。手法の 1) は実験の結果、有効性を確認した。手法の 2) の評価は今後の課題であるが、一方のメニューでの選択項目の分類語が、もう一方のメニュー体系ではどのように分布しているかを表示することで、シソーラスの項目名の文書での意味が特定しやすく、目次の各項目の記述内容が想定しやすくなると思われる。このシステムでは、二つのメニュー体系を並行して使用できるため、自分の直感により合ったわかりやすいメニューに任意の時点で移行できる利点もある。また、両体系を使った複合的な検索が可能のため、単独の場合より検索効率や適合率が良くなることが期待される。

参考文献

- [1] <http://www.cii.ipa.go.jp/nbdb/library1.html>.
- [2] Cutting Douglass R. Constant interaction-time scatter/gather browsing of very large document collection. In *Proceedings of ACM-SIGIR'93*, pp. 126-135, 1993.
- [3] 塩見, 他. シソーラスを用いた文書データの自動分類法. 情処研報 97-NL-117, pp. 99-104, 1997.
- [4] 国立国語研究所 (編). 『分類語彙表』. 国立国語研究所資料集 6. 秀英出版, 1964.
- [5] 千田恭子, 他. 汎用シソーラスを利用した検索用索引メニューの構成法. 情処研報 96-NL-111, pp. 21-26, 1996.
- [6] 日本経済新聞社. 日経全文記事データベース - 検索の手引 [第二版] -. 1984.