

分類体系と名詞句を用いた検索インタフェースの提案とその評価

石井 恵 渡辺 一成

NTT サイバースペース研究所

あらまし： 本稿では「おおまかな一覧／絞り込み」と「具体的な一覧／絞り込み」ができる検索インタフェースをもつ直観的な絞り込みが可能なロボット型検索システムを提案する。「おおまかな一覧／絞り込み」は分類体系上に検索結果をマッピングし、分類体系上のカテゴリを単位とした検索結果の絞り込みを可能とする。「具体的な一覧／絞り込み」は検索語を構成する単語を含む名詞句を提示し、それら名詞句を単位とした絞り込みを可能とする。本システムを自治体のホームページに適用した結果、検索結果の最初のページに表示される文書数以上の文書がヒットした際、ユーザは絞り込みの機能を利用する傾向があり、本機能の必要性が確認された。また、ユーザは分類体系を用いたおおまかな絞り込みより、名詞句を用いた具体的な絞り込みを利用する傾向があり、ユーザの検索支援における検索語の直接的かつ具体的な表現を用いた提示法の有用性が確認された。

Rough and Rigid Search Result Specialization Interface using Categorization Tree and Noun Phrase List

Megumi ISHII and Kazunari WATANABE

NTT Cyber Space Laboratories

Abstract: This paper proposes a new Web search engine that enables users to specialize search results intuitively. It has a new search interface to allow users to roughly or rigidly specialize search results and to flexibly combine those specializations. This interface uses a categorization tree for rough specialization and a list of noun phrases containing words included in the query terms for rigid specialization. We applied our system to a local government homepage. This trial showed that users need our specialization functions more when the number of documents that matched a query exceeds that of the documents displayed in the first page of the search result. Moreover, a new characteristic of the user search process is discovered; users prefer rigid specialization to rough specialization.

1 はじめに

インターネットロボットと呼ばれるプログラムにより Web 上の文書を収集し、ユーザにその検索を可能にする Web サーチャエンジン (AltaVista 等) は、インターネット上の情報取得になくてはならないツールとなっている。近年、Web サーチャエンジンが返却する文書数はますます増加し、ユーザが探している文書を Web サーチャエンジンの結果からみつけることは難しくなっている。その大きな原因のひとつは、Web 上の文書が爆発的に増加しているにもかかわらず、ユーザが入力する検索語の数が少ないことがあげられる。実際、検索システムにユーザが与える検索式を構成する検索語の数は非常に少なく、ほとんどは 1,2 個

である [1]。

検索式を検索語の集合としてのみではなく、検索語を論理結合した論理式として表すことを可能とする Web サーチャエンジンも多い。そのようなサーチャエンジンではユーザが最初から複雑な論理式を使って明確に表した検索式を与えることができれば、検索結果を少数に抑えることは可能であるかもしれない。しかし、[2]によると、与えられた課題を解くために必要な情報を Web 上で探す過程を調べる実験の中で、ユーザはもとの検索式に検索語を追加して絞り込むことが多いということが報告されている。このことは、ユーザに複雑な論理式を使って検索要求を記述させることは難しいことを意味すると解釈できる。よって Web サーチャエンジンから返却される文書集合の中か

らユーザが容易に目的とする文書を見つけだせるようにするためには、複雑な論理式を記述することなく、ユーザが直観的に検索結果を絞り込みこめることが必要となる。

そこで本稿では、直観的な絞り込みが可能な Web サーチエンジンを提案する。本システムでは、絞り込みに用いることによって得られる文書の特徴がユーザにとって理解しやすい表現を用いて、検索結果中に存在する検索語と関連する概念を表す。そして、それら概念を一覧性高くユーザに提示し、ユーザがその一覧から目的の概念を選択することにより、直観的な絞り込みを実現する。

本システムは、「おおまかな一覧／絞り込み」と「具体的な一覧／絞り込み」を可能とする新しい検索インタフェースをもつ。「おおまかな一覧／絞り込み」は、ユーザが検索結果中に存在する検索語に関連する抽象的な概念を一覧でき、抽象的な概念で検索結果の絞り込みを可能とする機能である。検索結果中の概念の関係は木構造を用いて表わす。本稿ではこの木構造を分類体系、木構造上のノードをカテゴリと呼ぶ。ユーザは分類体系上のカテゴリの選択により、そのカテゴリに対応する概念に関する特徴をもつ文書に検索結果を絞り込むことが可能である。

「具体的な一覧／絞り込み」は、検索結果中に存在する検索要求に関連する具体的な概念をユーザが一覧でき、その具体的な概念で検索結果の絞り込みを可能とする機能である。本システムは具体的な概念をユーザが入力した検索語を構成する単語を含む名詞句として表してユーザに提示する¹。ユーザは提示された名詞句の選択により、検索結果をその名詞句を含む文書集合に絞り込むことが可能である。本システムでは、ユーザにとってわかりやすい単位で荒く、細かく、両面からユーザが検索結果中の概念を一覧・絞りこめる点が従来のシステムと異なる。

以下、第2章では直観的な絞り込みの実現に対するアプローチについて述べる。第3章では本システムの構成を説明し、「おおまかな一覧／絞り込み法」と「具体的な一覧／絞り込み法」を提案する。第4章では本システムを自治体のデータに適用した結果を示す。第5章はまとめである。

2 検索結果の直観的な絞り込みの実現に対するアプローチ

2.1 検索結果中の概念の表現形式

直観的な絞り込みに必要となる、検索結果中の概念の提示のための従来の研究としては、大きくわけて出現する単語に着目したものと文書間の関係に着目したものがあ。単語に着目したアプローチとしては、検索式に追加する単語の提示を行なう **excite**², **RCAAU** [4] や、検索結果中の単語の出現頻度と文書内の共起関係

¹本稿ではユーザにより空白やカンマ等で区切られた各文字列を検索語と呼ぶ。検索語は1つ以上の単語から構成される。

²<http://www.excite.co.jp>

を用いて得られる単語間の関係を表すグラフの提示 [3] がある。**excite** では単語によってのみ概念が表されるため、絞り込みにその概念を用いた場合、得られる文書の特徴は得てして漠然としたものとなる。また **RCAAU**、[3] は単語間の関係を用いて概念を表すことができるが、検索結果から動的に利用する単語が決定されるため、必ずしもユーザがみてわかりやすい概念が生成されるとは限らない。その結果、提示された概念によって絞り込まれる特徴をユーザが想像しにくいという問題がある。

文書間の関係に着目したアプローチとしては、検索結果中の文書を概念に対応するいくつかのカテゴリに分け、そのカテゴリに属する文書の特徴を概念として提示する。文書の分け方としては、予め分類するカテゴリを定義して文書を分類する分類手法、検索結果から動的にカテゴリを生成するクラスタリング手法がある。分類手法を採用しているシステムの代表的な例は **Yahoo** があげられる。この手法では予め人によってカテゴリが定義されているため、ユーザはカテゴリに属する文書の特徴を理解しやすい。

クラスタリング手法を採用しているシステムとしては、[5], [6], [7] がある。クラスタリング手法では、予めカテゴリを定義する必要がないため、カテゴリのメンテナンスが不要という利点をもつ。反面、出現する単語に依存してカテゴリが生成されるため、必ずしも意味のあるカテゴリが生成されるとはかぎらないという問題をもつ。そのため、生成されたカテゴリが何を表しているのかをユーザが理解するのが時として困難となる。

2.2 本アプローチ

そこで我々は直観的な絞り込みを実現するため、絞り込みによって得られる文書の特徴を表す表現として、カテゴリをノードとした木構造で表わされる予め人が定義した分類体系のみでなく、検索結果の文書中に出現するユーザが入力した検索語を構成する単語を含む名詞句も用いる。両表現をもちいることにより、ユーザは検索結果中の概念を荒く、細かく、両面から一覧でき、一覧性高く眺めることが可能となる。更に、提示した名詞句、分類体系上のカテゴリでの検索結果の絞り込みをユーザにできるようにする。

ユーザが入力した検索語を構成する単語を含む名詞句は、ユーザが入力した検索語と関連のある概念を明示的に表すことができる。そのため、その概念が表すイメージをユーザにより容易に想像させることができる。例えば「高齢者 制度」が or 検索の検索語として入力された場合、名詞句である「高齢者居室等整備促進金融融資制度」を関連のある概念として提示する。名詞句による検索は、名詞句を構成する単語を and で結合して検索するより、単語の順序や各単語が共に出現する範囲を制限できるため選択された概念に対する適合率の高い検索が可能となることが期待できる。この名詞句を利用した検索を本稿では「具体的な一覧／絞り込み」と呼ぶ。

一方、人が定義した木構造で表わされる分類体系で

はカテゴリのラベルとして、そのカテゴリを表す語が人手によって通常1,2個割り当てられる。ラベルを構成する語の少なさから、カテゴリが表す概念はユーザにとって抽象的なものとなる。しかし、人間により意味づけられて作成されているので、体系自体は各カテゴリに属する文書の特徴を把握する手がかりとなる。体系上のカテゴリの選択による検索結果の絞り込みは、カテゴリのラベルで使用されている語を検索語とする検索に対し、選択した概念に対して再現率の高い検索を可能とすることが期待できる。これはラベルで用いられる語とは独立に、カテゴリには文書中に現れる単語を用いて予め似ている文書がまとめられているためである。この分類体系を利用して検索を本稿では「おおまかな一覧／絞り込み」と呼ぶ。

3 提案 web サーチエンジン

3.1 システム構成

本システムを図1に示す。本システムでは、Web上の文書の収集にTITAN [8]のロボットを、形態素解析プログラムとしてChaSenを用いる。また全文検索プログラムとしてWAIS-sfを用いる。本システムでは、形態素解析プログラムによって区切られた文字列を単語として扱う。

ロボットはもと文書及びそれを日本語形態素解析プログラム ChaSen を用いて単語分割を行なった文書を文書DBに格納する。単語分割された文書からWAIS-sfを用いて全文検索用のインデックスを作成し、インデックスDBに保存する。

同時に、本稿で提案する分類体系、名詞句による絞り込みがユーザの絞り込み要求に対して高速に行なえるよう、前処理として以下の処理を行なう。分類体系を用いた絞り込みのために、収集した文書を体系上のカテゴリへ自動分類し、収集した各文書からその文書が属するカテゴリを引くためのインデックスを分類DBとして作成する。名詞句による絞り込みのために、収集した文書から名詞句を自動抽出し、各文書からその文書が含む名詞句を引くためのインデックスと名詞句からその名詞句を含む文書を引くためのインデックスを名詞句DBとして作成する。検索エンジンはこれら情報を利用して、ブラウザからの検索要求を処理する。

3.2 ユーザインタフェース

本システムではユーザは検索式を検索語の集合で表現する。語の間の区切りは空白である。ユーザは入力した検索語の解釈の仕方として(a)全ての語を含む、(b)いずれかの語を含む、のいずれかを指定できる。

図2は検索式“高齢者 制度”(or 検索)に対する本システムの検索結果を示す。検索結果画面は、(1)検索式に適合した文書N件のタイトルを文書の適合順に表示する領域、(2)検索式に適合した文書から抽出された名詞句を表示する領域、(3)検索式に適合した文書が属するカテゴリと分類体系のルートからそのカテゴ

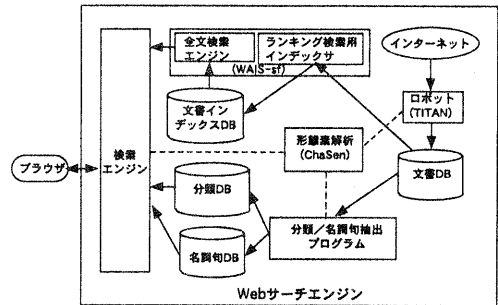


図1: システム構成図

りまでのパス上にあるカテゴリのみを表示した分類体系の部分木を表示する領域から構成される。

ユーザは、分類体系上のフォルダ形のアイコンの選択により、そのカテゴリに分類されている文書集合のリストを得ることができる。表示されている任意のフォルダの選択により、文書の特徴によって分けられた検索結果中の文書集合間の渡り歩きが可能である。名詞句は検索語を構成する単語の部分が赤で強調されて表示される。絞り込みに利用したい名詞句の左の丸の選択により、ユーザはその名詞句を含む文書の集合に検索結果を絞り込むことができる。「関連語句一覧」と書かれている左の丸の選択により、検索結果中に含まれる名詞句の一覧の取得が可能である(図3)。

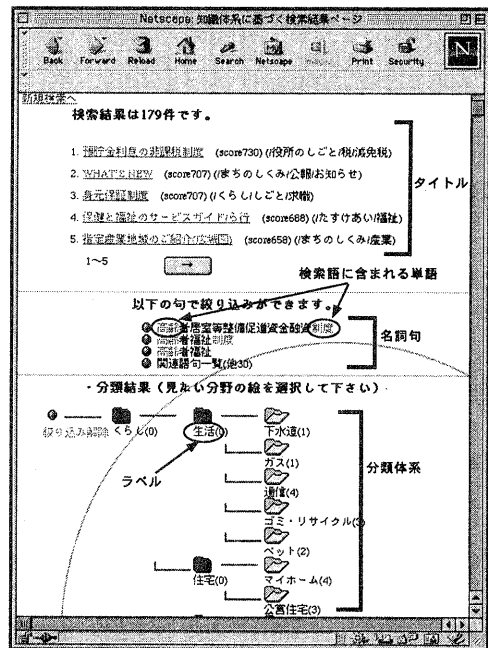


図2: “高齢者 制度”に対する初期検索結果画面

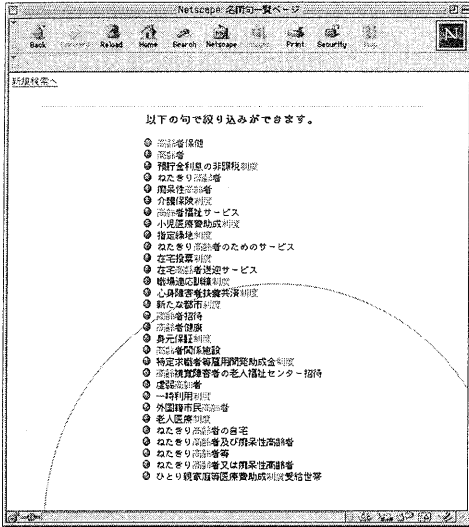


図 3: 名詞句一覧画面

3.3 おおまかな一覧／絞り込みのための文書分類

ロボットが収集した文書に対しておおまかな一覧／絞り込みを実現するためには、収集した文書を分類体系上に自動的に分類する必要がある。本システムではまず学習プロセスで体系を作成し、その後、収集した文書の分類を分類プロセスで行なう。

学習プロセス 人手で木構造をもつ分類体系を定義し、分類体系上にサンプル文書を手動で分類する。システムは分類された文書を形態素解析し、各カテゴリの特徴を各単語のそのカテゴリに対する貢献度で表すスコアテーブルを作成する（スコアテーブルの作成は手法 [9] に従う）。

分類プロセス 与えられた文書を形態素解析し、出現単語を求める。出現単語と学習プロセスで作成されたスコアテーブルから、その文書の各カテゴリに対するスコア値を求め、最大のスコア値をもつカテゴリへ文書を分類する（スコア値の計算は手法 [9] に従う）。不適切な分類に対しては、手動で分類先を変更し、再度スコアテーブルを作成しなおすことで対応する機構をもつ。

3.4 具体的な一覧／絞り込みのための名詞句抽出

前述のように分類体系上のラベルを構成する語の少なから生じるカテゴリの特徴の抽象性により、分類体系を用いた絞り込みは必ずしもユーザにとって容易とは限らない。更に分類体系による絞り込みは、必ずしもその分類体系がユーザに適したものであるとは限らないという欠点をもつ。これは絞り込みのための単位が、ユーザの検索式とは独立に予め固定させてしまっているからである。

そこでこれら欠点を克服するため、本システムでは検索結果中に存在する具体的な概念を提示する。本シ

ステムでは具体的な概念を検索語を構成する単語を含む名詞句により表す。前記名詞句は、ユーザが入力した検索語と関連のある概念を明示的に表すことができ、その概念が表すイメージをユーザにより容易に想像させることができるという利点をもつ。更に、本名詞句を用いた絞り込みでは、ユーザが入力した検索語から動的に絞り込まれる単位が決定されるため、ユーザの検索要求に適応した絞り込みが可能であるという利点をもつ。

以下に名詞句の自動抽出法について説明する。

3.4.1 名詞句の定義

本システムの名詞句を構成する要素を表 1 に示す。図 4 は名詞句を表すパターンを示す。名詞句パターン定義において“(と)”で囲まれる部分は省略可能、“+”は連続、“/”は「または」を意味する。収集された文書に対して形態素解析を行ない、単語とその単語の品詞情報を取得する。定義したパターンにマッチする品詞の並びをもつ単語列を名詞句として抽出する。この時、単語列はパターンとの最長マッチで切り出される。

3.4.2 提示名詞句の選出

本システムでは予めユーザへ提示する名詞句の最大数 N を設定しておく。検索語の入力に対して提示される名詞句の定義は以下の通りである。

【提示名詞句の定義】 検索結果に含まれる文書の中に存在し、かつ、ユーザが入力した検索語に含まれる単語を含む名詞句。 □

提示名詞句は以下の優先順序により、検索結果の画面に上位 N 個をユーザに提示する。

【提示戦略】

- (1) ユーザが入力した検索語に含まれる単語を多く含む名詞句ほど上位に提示。
- (2) 検索語に含まれる単語を含む数が同じ場合、文書内で出現頻度が高い名詞句ほど上位に提示。

上記戦略は以下の思想にもとづく。

- ・ ユーザが入力したより多くの検索語で表される概念ほど、ユーザが探している概念である。
- ・ ユーザはある名詞句で絞り込んだ場合、当然その名詞句で表される概念を主題とした文書の集合を検索結果として期待する。

名詞句の絞り込みでは、1つの文書から複数の名詞句が同時に提示される場合がある。これは名詞句を1つのカテゴリとみなすと複数のカテゴリへの文書の分類を許すことを意味する。文書は複数の主題をもつ場合があり、その場合、複数のカテゴリに分類できることが望まれる。本システムの名詞句提示戦略による名詞句による絞り込みではその要求を満たしている。

表 1: 名詞句を構成する品詞

品詞	例
名詞	-
名詞接続助詞	の、かつ、また等
名詞接頭辞	お〇〇、ご〇〇等
名詞性名詞接尾辞	〇〇用、〇〇者、〇〇以下等
名詞性名詞助数詞	〇〇回、〇〇人、〇〇m 等
形容詞	-
形容詞性名詞接尾辞	〇〇的な、〇〇風な等

句	= (修飾部) + 名詞部
修飾部	= 形容詞部 + 句 / 名詞部 + 名詞接続助詞 + 句 / 名詞部 + 句
名詞部	= (名詞接頭辞) + 名詞 + (名詞性名詞接続辞 / 名詞性名詞助数詞)
形容詞部	= 形容詞 / 名詞 + 形容詞性名詞接続辞

図 4: 名詞句パターン定義

一方、本名詞句提示戦略は、名詞句の絞り込みによって検索結果中の文書全に必ずしも到達できることを保証はしない。すなわち検索結果中の全ての文書の内容を名詞句によって一覧できるとは限らないという一覧性の不完全性をもつ。しかし本システムでは、全ての文書は必ず分類体系上のいづれかのカテゴリに分類されることを保証しているため、分類体系による一覧によりその一覧性の不完全性を補うことができている。

4 評価結果

本システムをある自治体のホームページのコンテンツに対して適用したサービス実験の結果を述べる。具体的には、(1)自動分類の精度、(2)抽出された名詞句の性質、(3)ユーザの絞り込み機能の利用状況について述べる。分類体系は5箇所の自治体のホームページの構成と市役所が発行している市民の生活ガイドである印刷物の構成を参考に作成した。分類体系の大きさはカテゴリ数79で、深さ最大5段、平均約4段である。

実験システムではインターネット上で公開するとともに、街頭端末(KIOSK)からのアクセスも可能にした。検索語の入力のインターフェースとして、通常のキーボードを用いたテキスト入力他に、キーボード入力ができないKIOSK用としてメニューからの検索語の選択により検索語の入力ができるインターフェースを用意した。

4.1 自動分類精度

本実験では853の文書をテスト用文書100、残りを分類体系のカテゴリの特徴を学習するための訓練文書に分けた文書セットを3セット作成し、分類の精度の

平均を調べた。各文書の特徴は各文書全体を用いて抽出した。前記抽出された特徴から獲得した分類体系のテスト用文書に対する精度は平均42%であった。

分類精度を向上させるためには、文書の特徴をより的確に表す必要がある。そこでタイトルのみを用いて文書の特徴を抽出した。これはタイトルは文書の内容を比較的表していると考えたためである。その結果、獲得された分類体系のテスト用文書に対する精度は平均53%となり、精度の向上を図ることができた。なお、タイトルのみを用いて文書の特徴を抽出した場合の、853全文書を訓練事例として与えて獲得した分類体系の訓練事例そのものに対する分類精度は81%である。

4.2 抽出名詞句の特徴

分類の実験に利用したコンテンツに新たな文書を追加して本システムのサービス実験を行なった。文書数は1348である。ユーザに提示する名詞句の数の上限は、図2の検索結果については3、図3の画面に表示される名詞句数は最大30とした。よって、ユーザは最大33個の名詞句をみることができる。

名詞句抽出の正確性

本実験では25047の名詞句が抽出された。抽出された名詞句のうち、数字の連続、単語の切り出しミス等、人が見て不自然な切り出し方をされた文字列(非名詞句文字列)は18.4%であった。このうち文字列の中にユーザにより検索語として入力されるおそれがあるもの(普通名詞、固有名詞を含むもの)は、37%であった。しかし、これらが抽出された文字列全体に占める割合は6.8%と少なくサービス提供にあたって大きな問題とはならないと考えられる。しかし、インデックス作成の処理を軽減するためにも非名詞句文字列を削除するパターンを考えることが今後必要である。

抽出された名詞句の具体性

ユーザが入力した検索式を構成する検索語は平均1.34であり、当初述べたように少ない。ユーザが入力した検索語を構成する平均単語数は1.3単語であった。一方、検索結果に提示される候補である、実際ユーザが入力した検索語を構成する単語を含む名詞句を構成する単語数は平均3.6であった。よって検索語は約2.8倍に拡張されている。これにより、入力された検索語と関連する具体的な一覧が生成されること、また名詞句を拡張検索語として利用可能なことがわかる。

提示名詞句の文書の特徴記述力

文書内で多く出現する文字列は、その文書の特徴を表す傾向があると考えられる。タイトルに現れる名詞句の文書内の最大出現回数の平均は2.27である。一方、タイトルに現れない名詞句の文書内の最大出現回

表 2: 絞り込み種別とその実行回数

名詞句のみによる絞り込み	364 回
分類体系のみによる絞り込み	235 回
両者を用いた絞り込み	58 回

表 3: 検索結果件数と絞り込み利用率の関係

検索結果件数	検索発生数	絞り込み発生数	利用率
1 - 5	492	87	0.18
5 - 50	1061	328	0.31
50 - 100	407	132	0.32
100 - 150	97	33	0.34
150 以上	193	77	0.39

数の平均は 1.25 回である。本名詞句の提示戦略では、名詞句内にユーザが入力した検索語に含まれる単語が同数含まれる場合は、文書内最大出現回数の多いものを優先する。よって、本名詞句の提示戦略はタイトルに出現する名詞句を優先する働きがあり、提示される名詞句は文書の特徴を比較的表していると考えられる。

4.3 ユーザによる絞り込みの特徴

表 2 は前記実験でユーザによって行なわれた絞り込みの様子を示す。検索の単位は、絞り込みが行なわれない検索から絞り込み行なわれない次の検索までを 1 つとして検索ログから切り出した。よって 1 つの検索の中で絞り込みは複数回行なわれることがある。

全検索数 2057 の 20 % である 411 の検索において絞り込み機能は利用された。そして、その中では句による絞り込みの方が分類による絞り込みより多く利用されている (表 2)。すなわち、句による絞り込み機能の方がユーザに好まれて利用されている。このことは、ユーザにとっては大雑把に全体を一覧できるよりも、検索語に関連するより具体的な概念を提示された方が、利用しやすいということを意味する。

検索結果件数と絞り込みの関係を表 3 に示す。検索結果件数とはユーザが入力した検索語にヒットした文書数である。一画面に表示される検索結果の文書数は 5 件とした。表中の「利用率」とは絞り込み発生数を検索発生数で割ったものである。検索結果の文書数が検索結果の最初の画面に表示される文書数よりも多い場合、明らかに絞り込み機能が利用される割合が高くなっている。これは見つかった文書が検索結果画面に収まらない場合に本絞り込み機能が特にユーザに必要であることを示していると考えられる

5 まとめ

本稿では直観的な絞り込みが可能な Web サーチェンエンジンを提案した。本システムでは分類体系を用いた「おおまかな一覧/絞り込み」と名詞句を用いた「具体的な一覧/絞り込み」により、直観的な絞り込みを実現している。一般的に全ての分野をカバーする分類

体系を作成することは多大なる労力を要する。そこで行政、教育といったように分野を限定して分類体系を作成し、同一分野内の複数の web サイトを横断的に検索するシステムとして本システムを利用することが適切であると考えている。今後の課題としては (1) 間違った分類は、ユーザの検索効率を下げるので、分類精度の向上、(2) 大量の文書へも適用するための、名詞句のインデックス作成の高速化、(3) 多量に名詞句が抽出された際のユーザにわかりやすい名詞句の提示法が必要である。

参考文献

- [1] P.G.Anick: Adapting a Full-text Information Retrieval System to the Computer Troubleshooting Domain, In Proc. of ACM/SIGIR'94, pp.349-358 (1994).
- [2] 白澤、新垣、野島、石崎: WWW 検索行動における「戻る」行動と検索方針の変化との関係、情処研報, HI83-11, pp.61-66 (1999).
- [3] 丹羽: 動的な共起解析を用いた対話的文書検索支援、情処研報, NL115-14, pp.99-106 (1996).
- [4] 西村、伊藤、河野、長谷川: 重み付き相関ルール導出アルゴリズムによる WWW データ資源の発見, DEWS'96, pp.79-84 (1996).
- [5] O.Zamir, O.Etzion: Grouper: a dynamic clustering interface to Web search results, In Proc. of the 8th International World Wide Web Conference, pp.283-296, Elsevier Science, (1999).
- [6] M.A.Hearst and J.O.Pedersen: Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results, In Proc. of the ACM/SIGIR'96, pp.76-84 (1996).
- [7] P.G.Anick and S.Vaithyanathan: Exploiting Clustering and Phrases for Context-Based Information Retrieval, In Proc. of the ACM/SIGIR'97, pp.314-320 (1997).
- [8] Y.Hayashi, G.Kikui and S.Suzaki: TITAN: A Cross-linguistic Search Engine for the WWW, AAAI Spring Symposium Technical Report, SS-97-05, pp.56-62 (1997).
- [9] 田村、渡辺、原、笠原: 統計的手法による文書自動分類, 第 36 回情処全大, 6U-5, pp.1305-1306 (1988).