

階層分類されたコンテンツ検索のための キーワードナビゲーションの提案

岡城 純孝[†] 河野 泉[†] 宮崎 陽司[‡]

[†] NEC 共通基盤ソフトウェア研究所

[‡] NEC サービスプラットフォーム研究所

階層分類されたコンテンツ検索では「次に選択すべきディレクトリをユーザが判断できない」「類似コンテンツが複数の異なる項目に分類されてしまう」などの問題がある。本稿では、1つの分類項目内のコンテンツ集合に特徴的なキーワードである特徴キーワードと、複数の分類項目に跨って存在するキーワードである共通キーワードを、ディレクトリ型検索における選択候補ディレクトリ一覧などと組み合わせて提示することにより、ユーザの検索を支援するキーワードナビゲーションを提案する。

Proposal of Keyword Navigation for Contents Search Based on Directory Structure

Sumitaka OKAJO[†] Izumi KOHNO[†] and Yoji MIYAZAKI[‡]

[†] Common Platform Software Research Laboratories, NEC Corporation

[‡] Service Platforms Research Laboratories, NEC Corporation

We propose a navigation method for contents search based on directory structures, focusing on keyword presentation-based methods. This proposed method uses “feature keywords” and “shortcut keywords”. Feature keywords are terms characterizing a specific directory, appear in contents in the directory. Shortcut keywords are terms distributed in contents in sibling directories. We also propose a search interface that effectively presents those keywords to searching users.

1. はじめに

従来から FAQ や商品情報など多くの種類のコンテンツが階層構造（ディレクトリ）で管理されている。階層構造は、ひとかたまりの意味を持つコンテンツ集合が何らかの分類基準によってディレクトリに分類され、さらに上位層から下位層にいくに従って分類基準が細分化されるようにして形成される。通常、このように管理されたコンテンツを検索するユーザは、各ディレクトリに付けられたディレクトリ名を参照してディレクトリを選択しながら上位層から下位層へ階層構造を辿ることによってコンテンツを絞り込んでいき、最終的に目的のコンテンツに到達する。

このとき、ユーザが適切でないディレクトリ

を選択してしまうと、目的のコンテンツに到達するために階層を上下に何度も辿らなければならず、検索効率が悪くなる。

本稿では、階層分類されたコンテンツを検索するユーザを支援するために、1つのディレクトリに分類されたコンテンツ集合に特徴的なキーワードである特徴キーワードと、複数のディレクトリに跨って存在するキーワードである共通キーワードを提示するキーワードナビゲーションを提案する。

2. ディレクトリ型検索

“階層分類されたコンテンツ”とは、コンテンツをディレクトリと呼ぶ任意のまとまりに分類して格納し、ディレクトリに上位下位の階

層関係を持たせた構造で分類されたコンテンツを指す。上位ディレクトリの中には下位ディレクトリが格納されており、下位ディレクトリの中には個々のコンテンツか、あるいはさらに下位に位置するディレクトリが格納されている。この構造によって多数かつ多様なコンテンツを意味や用途によって分類することができる。このような階層構造コンテンツのメリットは、その分類構造が人間にとって直感的であり、わかりやすいことが挙げられる。また、構造全体を明示すれば、分類の全体像を把握しやすいこともメリットである。

階層分類されたコンテンツを検索するユーザは、通常、一覧表示される同階層のディレクトリから一つを選択し、さらに続いて一覧表示される、選択ディレクトリの一段下位層のディレクトリから一つを選択することを繰り返し、階層構造を上位層から下位層へ辿ることによってコンテンツを絞り込み、最終的に目的のコンテンツに到達する。ユーザは、選択肢として表示されるディレクトリを選ぶだけで検索結果を絞り込みながら目的のコンテンツに到達することができる。

3. 関連研究

宮本らは、情報を絞り込まず、長い検索結果リストを大雑把に読み飛ばしつつも必要な情報を見逃さないような検索結果の表示方法を提案している[1]。提案方法では、共通あるいは共起頻度の高い単語を多く含む Q&A が隣接してリスト表示される。これによってユーザは検索結果の Q&A を一つずつではなく、ある程度まとめて検索要求に該当しそうか否かを判断できる。リスト内で意味のまとまりで分類されており、階層構造を遷移するよりもシームレスであることを特徴としている。しかし、この方法は、リストから目的のキーワードが隣接している部分を探す必要がある。

Right Now[2]は、過去の利用履歴を解析し、検索されたコンテンツの参照後によく利用されるコンテンツを自動的に表示することによって、参照したコンテンツがユーザの目的とするものと異なる場合、他のコンテンツへのアクセスや関連発見を促している。しかし、この方法は、ユーザの検索過程におけるディレクトリ選択については何ら支援することができない。

また、あらかじめ登録しておいた、検索に必要な語句や専門用語の一覧をディレクトリと

ともに提示するものもある[3]。ユーザは、それらのキーワードを選択することで正しい語句で検索を実行できるようになる。しかし、この方法はあらかじめキーワードを登録する必要がある。

我々は、ユーザが選択肢として提示されるキーワードを選んでいくだけで目的のコンテンツに到達できるという点で、3つ目の方式に着目した。しかし、前述の方法では、ディレクトリ検索とキーワード選択による検索が互いに独立しており、ユーザが目的のコンテンツに到達するための支援をうまく行えているとは言えない。つまり、あらかじめ登録したキーワード一覧を提示するだけであるので、ディレクトリ検索で絞り込んだ検索結果とキーワードとの関連性がなく、ディレクトリ探索途中でキーワードを選択しても検索結果が0件になってしまうなどコンテンツをうまく絞り込めないことが多い。そのため、階層分類されたコンテンツに対してディレクトリ検索を行うユーザを目的のコンテンツへ効率よくナビゲートする技術が必要である。

ディレクトリとキーワードの関係について谷津らは、検索キーワードが Web ディレクトリのどの位置に属するかを判定する方法を提案している[4]。Web ディレクトリ上のあるカテゴリ名と、そこに属するサイトの照会文から単語を抽出し、機械学習により Web ディレクトリという体系における単語の位置を推定している。また、北内らは、語と文書の関係を用いた重み付けである *tf-idf* と、語とカテゴリの関係を用いた重み付けを組み合わせることで、分野（カテゴリ）情報を利用した語の重み付け手法を提案している[5]。

4. ディレクトリ型検索における課題

ディレクトリ型検索が有効に機能するためには、階層構造を形成するための分類基準がしっかり整理されていることが重要となる。つまり、あるコンテンツがどのディレクトリに分類されるか、ほぼ一意に決まるということが重要であり、そうでなければ後でディレクトリを辿ってコンテンツを検索することができなくなってしまう。しかしながら、新たなコンテンツを登録する際などに、当初構成した分類基準に当てはまらないコンテンツや、複数の分類基準に当てはまるコンテンツが出てくることも多い。

また、階層構造や分類基準は、コンテンツ管理者側で作成されるため、ユーザが想定する分類基準と合わない場合がある。このような場合、ユーザはディレクトリ名から分類基準を判断しなければならず、次に選択して辿るべきディレクトリがわからなくなるという問題がある。また、専門用語など、そもそもディレクトリ名の意味内容が理解しにくい場合には分類基準がわからないために、ユーザが次に選択すべきディレクトリがわからなくなるという問題もある。さらに、ユーザが想定する分類基準では同一のディレクトリに存在するような内容の複数のコンテンツが、異なる複数のディレクトリに分類されてしまうという問題もある。

以上、まとめると階層構造コンテンツのディレクトリ型検索には、ユーザの検索効率を損なう原因となる、次のような課題があると考えられる。

【課題1】

コンテンツ管理者側であらかじめ定義される分類基準がわかりにくい、ディレクトリ名がわかりにくい、あるいは検索ユーザが想定する分類基準と合わない場合には、ユーザは次にどのディレクトリを選択すれば良いかわからない。

【課題2】

コンテンツ管理者側であらかじめ定義される分類基準によって、近い内容のコンテンツが複数のディレクトリに分類されることがあり、この場合、ユーザは階層を上位に遡ってディレクトリを選択し直すなど複数のディレクトリをあれこれ選択しなければならない(図1参照)。

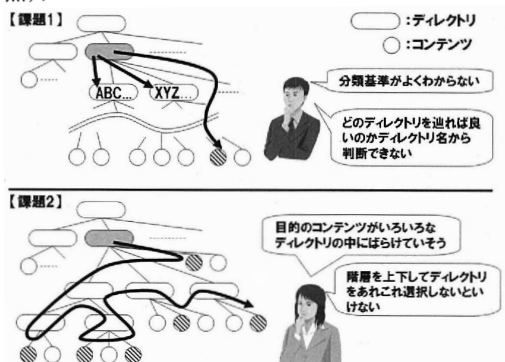


図1 ディレクトリ型検索における課題

5. キーワードナビゲーション

前節で述べた2つの課題を解決するために、検索対象コンテンツから抽出した2種類のナビゲーションキーワード(特徴キーワード、共通キーワード)を、ディレクトリ型検索における選択候補ディレクトリ一覧などと組み合わせ提示することにより検索ユーザの支援を行うナビゲーション機能(キーワードナビゲーション)を提案する。検索ユーザは、これらのキーワードを参照したり選択したりすることにより、目的のコンテンツを効率的に検索することが可能となる。

以下、特徴キーワードと共通キーワードについて詳細に説明する。

5.1. 特徴キーワード

特徴キーワードは、「ディレクトリ内のコンテンツに特徴的なキーワード」を意味し、前節の課題1を解決することを目的とする。

ディレクトリ型検索において、同一階層に複数のディレクトリが存在する場合に、ユーザが目的のコンテンツに到達するために次にどのディレクトリを選択すべきかをディレクトリ名から判断するのは非常に難しい。また、各ディレクトリのコンテンツ一覧を提示されたとしても、ユーザがコンテンツの内容を一つ一つ確かめなければならない。このような場合に、例えば、ディレクトリごとに出現頻度の高いキーワードを特徴キーワードとして抽出し提示することにより、ユーザは各ディレクトリ内のコンテンツ集合の概略を把握でき、適切なディレクトリを迷うことなく選択できる。

今、図2に示すような階層構造を持つコンテンツ集合を考える。ルートディレクトリrootは、2つのディレクトリAとBを持ち、ディレクトリAは3つのコンテンツA-1, A-2, A-3を持つ。ディレクトリBは2つのディレクトリCとDを持ち、ディレクトリCは2つのコンテンツC-1, C-2を、ディレクトリDは3つのコンテンツD-1, D-2, D-3を持つ。またディレクトリBはその直下に2つのコンテンツB-1, B-2を持つ。

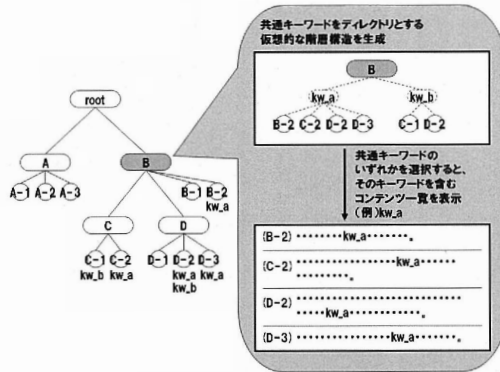


図 2 特徴キーワードの生成

ユーザがディレクトリ root に位置し、次に A と B どちらのディレクトリを選択すべきかをナビゲートするために、ディレクトリ A とディレクトリ B のそれぞれの特徴キーワードを生成する。このとき、ディレクトリ B 内には、ディレクトリ C およびディレクトリ D が存在するが、ディレクトリ B の特徴キーワードの生成では、その構造によらず、ディレクトリ B 以下のすべてのコンテンツ集合から特徴キーワードを生成する。言い換えると、ディレクトリ A についてはコンテンツ集合 {A-1, A-2, A-3} から、ディレクトリ B についてはコンテンツ集合 {B-1, B-2, C-1, C-2, D-1, D-2, D-3} からそれぞれ特徴キーワードを生成する。

このようにして生成した各ディレクトリの特徴キーワードをユーザに提示すると、ユーザは各ディレクトリ内のコンテンツ集合の概略を把握することができるようになる。つまり、ユーザはディレクトリ型検索中に目的のコンテンツに到達するために選択すべきディレクトリを判断することが可能になる。したがって、ユーザはディレクトリの選択に迷ったり、選択を誤ったりすることなく、最短ルートでディレクトリを辿って目的のコンテンツに到達することができ、コンテンツ検索を効率的に行うことが可能となる。

また、ユーザがディレクトリ B を選択して一階層下がった場合には、同様にして {B-1, B-2}, {C-1, C-2}, {D-1, D-2, D-3} のコンテンツ集合それぞれから特徴キーワードを生成し、ユーザに提示する。

5.2. 共通キーワード

共通キーワードは、「複数ディレクトリに跨って存在するキーワード」を意味し、前節の課題 2 を解決することを目的とする。

ディレクトリ型検索において、分類基準によってはユーザの目的に近い内容のコンテンツが一つのディレクトリにまとまって存在せず、複数のディレクトリに散在してしまう場合がある。

このような場合に、複数のディレクトリに跨って存在するキーワードを共通キーワードとして抽出し、新しいディレクトリとして提示することにより、ユーザはコンテンツ管理者側であらかじめ定義された分類によらず、キーワードに基づいた分類によってコンテンツを検索することができる。

今、図 3 に示すように、コンテンツ B-2, C-2, D-2, D-3 がキーワード "kw_a" を含み、コンテンツ C-1, D-2 がキーワード "kw_b" を含むとする。また、ユーザがディレクトリ B に位置しているとする。このとき、既存のディレクトリ C とディレクトリ D に加え、これらキーワード "kw_a", "kw_b" を含むコンテンツをまとめた仮想的なディレクトリを生成する。

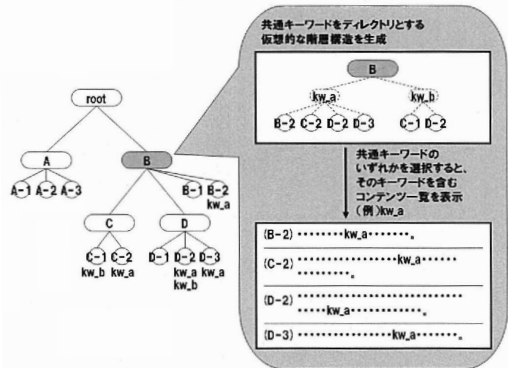


図 3 共通キーワードの生成

このようにして生成した共通キーワードを提示し、ユーザがそれを選択すると、選択したキーワードを含むコンテンツ一覧を表示することで、複数のディレクトリに散在するコンテンツをまとめて検索することができるようになる。例えば、キーワード "kw_a" に関連するコンテンツを探しているユーザは、キーワード "kw_a" を含むコンテンツが複数のディレクト

りに散在するために特徴キーワードを参照してもディレクトリを選択できない場合でも、"kw_a"を含むコンテンツ一覧を得ることができる。共通キーワードによって、ユーザは目的のコンテンツを検索するために、階層を上位に遡ってディレクトリを選択し直すなど複数のディレクトリをあれこれ選択する必要がなく、複数のディレクトリに散在するコンテンツにまとめて到達することができ、コンテンツ検索を効率的に行うことが可能となる。

6. 試作システム

階層分類された家電製品の FAQ を検索対象コンテンツとして、キーワードナビゲーションシステムを試作した。本節では、試作したシステムについて述べる。

6.1. システム概要

試作したシステムは、ディレクトリを辿ることによって目的の Q&A を検索するディレクトリ型検索が可能な FAQ 検索システムに、特徴キーワードと共通キーワードによるキーワードナビゲーション機能を追加したものとなっている。

図 4 に試作システムの構成図を示す。試作システムは、大きく分けてデータ生成部とナビゲーション部の 2 つから構成される。

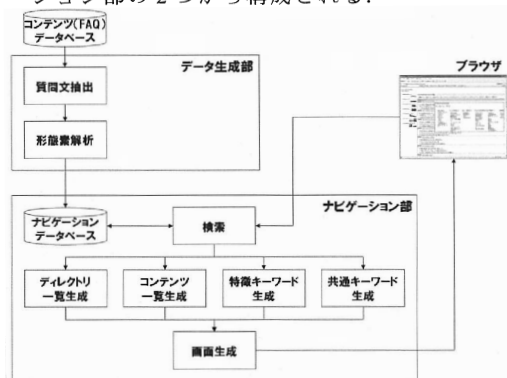


図 4 システム構成

データ生成部は、検索対象コンテンツからナビゲーションに必要なキーワードを抽出する処理を行う。質問文抽出部は、あらかじめ階層分類されたコンテンツが格納されているコンテンツデータベースから質問文を抽出する。続いて抽出した質問文を形態素解析しキー

ワードを抽出する。検索対象のコンテンツが FAQ であるため質問文だけでなく回答文を利用することも考えられるが、FAQ 検索においてユーザが検索クエリとして思い浮かべるのは質問文であると考え、今回は質問文からキーワードを抽出することとした。また、家電製品の FAQ に対する検索クエリとしては、機能や部品などの名詞が重要となると考え、名詞のみを抽出することとした。

ナビゲーション部は、ユーザがブラウザを介して選択したディレクトリやキーワードに対する検索の実行と、その検索結果に応じたナビゲーションキーワードの生成などを行う。ナビゲーションデータベースには、前述のデータ生成部で抽出したキーワードを、その抽出元のコンテンツ、およびそのコンテンツが属するディレクトリと関係付けて格納する。検索部は、ユーザが選択したディレクトリやキーワードに対してナビゲーションデータベースを検索する。ディレクトリ一覧生成部は、ユーザが選択したディレクトリに属する一階層下のディレクトリ一覧を生成する。コンテンツ一覧生成部は、ユーザがディレクトリを選択した場合には、そのディレクトリ、およびそのディレクトリの下位層のディレクトリに属するコンテンツ一覧を生成し、ユーザがキーワードを選択した場合には、そのキーワードの抽出元のコンテンツ一覧を生成する。特徴キーワード生成部および共通キーワード生成部は、ナビゲーションデータベースを参照し、検索結果のコンテンツ集合に対する特徴キーワードと共通キーワードを生成する（詳細は後述）。画面生成部は、ディレクトリ一覧生成部、コンテンツ一覧生成部、特徴キーワード生成部、共通キーワード生成部の出力結果を用いて検索結果画面を生成する。

試作システムの検索画面例を図 5 に示す。画面左部には、階層構造の大分類、つまりルートディレクトリ直下のディレクトリ一覧を表示している。これは、例えば代表的なディレクトリ型検索である「Yahoo!カテゴリ」[6]の TOP 画面と同様のものであり、ユーザはこの大分類のディレクトリを選択することから検索を開始する。図 5 では、製品種別が大分類となっている。画面右下部には、絞り込み途中の検索結果として、ディレクトリやコンテンツを表示している。図 5 は、画面左部の大分類から「DVD レコーダー」を選択した画面を示しており、ディレクトリ「DVD レコーダー」の下位層のディ

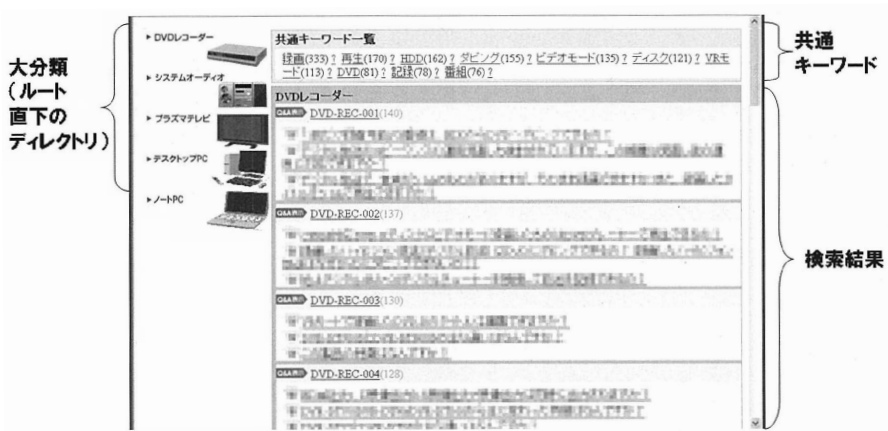


図 5 検索画面例

レクトリ一覧として、「DVD-REC-001」, 「DVD-REC-002」, 「DVD-REC-003」などを表示している。また、各ディレクトリに属するQ&Aコンテンツの一部も同時に表示している（機密情報のため、該当部分にモザイク処理を施している）。この後、ユーザが、表示されているディレクトリ一覧からディレクトリを選択すると画面右部を再描画し、選択したディレクトリに属する下位層のディレクトリ一覧と各ディレクトリに属するQ&Aコンテンツを表示する。ユーザは、この操作を繰り返してディレクトリを辿ってコンテンツを絞り込みながら FAQ 検索を行う。

6.2. 特徴キーワード表示

次に、特徴キーワードの表示画面例を図 6 に示す。試作システムでは、表示されているディレクトリ名にマウスオーバーすると、そのディレクトリの特徴キーワードをポップアップ表示する。図 6 は、ディレクトリ「DVD-REC-001」にマウスオーバーした画面を示している。



図 6 特徴キーワードの表示画面例

図 6 のポップアップ画面には、「DVD-REC-001」に属する下位層のディレクトリである「ディスク関連」、「ダビング関連」、「デジタル放送関連」などと、各ディレクトリそれぞれの特徴キーワードを表示している。特徴キーワードの数が多くなると見づらくなるため、このように、マウスオーバーしたディレクトリ直下のディレクトリごとに整理して表示している。なお、試作システムでは、各ディレクトリに属するコンテンツから抽出したキーワードを頻度でランキングし、一定の順位以上のものを特徴キーワードとした。

ユーザは、検索途中で表示される同一階層のディレクトリ名に順にマウスオーバーして各ディレクトリの特徴キーワードを参照することにより、それぞれのディレクトリに属するQ&Aの概要を把握することができ、目的のQ&Aに到達するために次にどのディレクトリを辿れば良いかを判断することができる。

6.3. 共通キーワード表示

次に、共通キーワードの表示画面例を図 7 に示す。試作システムでは、ユーザが選択したディレクトリに属するすべてのQ&Aコンテンツについて生成した共通キーワードを検索画面右上部に表示する。



図 7 共通キーワードの表示画面例

図 7は、まず大分類から「DVDレコーダー」を選択し、その下位層のディレクトリー一覧から「DVD-REC-001」を選択したユーザが、「再生」に関する Q&A を検索したいと考えた場面で、「再生」に関連するディレクトリが存在せず、さらに「再生」というキーワードが「ディスク関連」をはじめとする複数のディレクトリのコンテンツに散在しているときに、共通キーワード「再生」を選択した様子を示している。図 7で、「DVDレコーダー」選択直後と、「DVD-REC-001」選択直後で、一覧されている共通キーワードの内容が変化していることがわかる。これは、それぞれ選択したディレクトリの下位層のディレクトリ構造が異なり、ディレクトリ間に跨って存在するキーワードの数や種類が変化するためである。このように、共通キーワード一覧には、ユーザが現在選択中のディレクトリに対する共通キーワードが表示される。

ユーザは、共通キーワードを選択することにより、同一のキーワードを持ち、複数のディレクトリに散在するコンテンツに容易に到達することができる。

7. おわりに

本稿では、階層分類されたコンテンツを検索するユーザを支援するために、1つのディレクトリに分類されたコンテンツ集合に特徴的なキーワードである特徴キーワードと、複数のディレクトリに跨って存在するキーワードである共通キーワードを提示するキーワードナビゲーションを提案した。

特徴キーワードは、ユーザが各ディレクトリ内のコンテンツ集合の概略を把握することを可能にし、ディレクトリ型検索中に目的のコンテンツに到達するために選択すべきディレクトリの判断を支援する。

共通キーワードは、複数のディレクトリに散在するコンテンツをまとめて検索することを可能にし、ディレクトリ型検索中に階層を上位に遡ってディレクトリを選択し直すなど複数のディレクトリをあれこれ選択せず複数のディレクトリに散在するコンテンツ検索を支援する。

また、階層分類された家電製品の FAQ を検索対象コンテンツとしたキーワードナビゲーションの試作システムについても述べた。

今後は、提案方式の有効性や検索効率化への効果の評価を行う。また、キーワードの抽出方法やランキング方法、および検索インタフェースの改良・強化を行っていく予定である。

参考文献

- [1] 宮本勝, 中谷挽子, 渡辺昌洋, 米村俊一, 小川克彦, “意味のまとまりを考慮した FAQ の検索結果リストの表示方法,” 電子情報通信学会技術研究報告 HIP, ヒューマン情報処理, Vol.104, No.169, pp.7-12, July.2004.
- [2] <http://rightnow.mki.jp/>
- [3] <http://www.justsystems.com/jp/km/cbfaq/cbfaq01.html>
- [4] 谷津哲平, 新納浩幸, 佐々木稔, “Web ディレクトリを用いた検索ナビゲーション,” 言語処理学会第 11 回年次大会, pp.1022-1025, March.2005.
- [5] 北内啓, 小西一也, 高木徹., “情報検索における分野情報を用いた語の重み付け, DEWS2004, 2-B-5, March.2004.
- [6] <http://dir.yahoo.co.jp/>