

ノードによる Web 検索インタフェース

Web Information Retrieval Interface Using Nodes

奥村 穂高[†]

Hotaka OKUMURA

田中 二郎^{††}

Jiro TANAKA

[†]筑波大学 大学院修士課程 理工学研究科

Master's program in Science and Engineering, Univ. of Tsukuba

^{††}筑波大学 電子・情報工学系

Institute of Information Sciences and Electronics, Univ. of Tsukuba

概要

我々は、キーワードを表現するノードを操作し対話的に検索条件の指定を行うキーワード検索システムを作成した。本システムは、ユーザがポインティングデバイスを使いノードをドラッグ&ドロップすることでキーワードの取捨選択やクエリーの構成をおこないながら次第に Web ページを絞り込んで行くことを可能にする。検索結果は通常の検索エンジンが URL を提示するのに対して、Web ページより抽出したキーワード群によって表示する。これにより、ユーザのキーワード選択の負担を軽減する。

1 はじめに

現在、Web 検索を行う手法としてキーワードとなるテキストをキーボードやコピー&ペーストで入力することによって与えるものが普通である。このため、目的のページに合ったキーワードが思いつかない場合、検索は困難なものとなる。また、的確ではないキーワードを選定してしまうと、目的のものとは異なるページが多く引っかかってしまう。絞り込むための条件として更新時刻などの付加情報を利用するものもあるが、それらは操作が複雑になるため、あまり利用されていないのが現状である。

我々はこの問題の解決のために、従来のシステムのように URL を表示せず、ノード化したキーワードによってクエリーを作成し、提示されるキーワードによって目的のページを対話的に探し出す検索システムを作成している。このシステムでは、結果をキーワード群で提示する。このキーワード群をクエリーに利用することによって、対話的に検索を進め絞り込みを行うことができる。これは、重要なキー

ワードがわからないときなどに有効であり、煩雑なキーワードの策定によるユーザの負担を削減する。

2 ノードを用いた Web 検索

我々が提案する対話的な Web のキーワード検索システムの概要について述べる。

システムの画面は、主に左上部のクエリーの作成を行うクエリー作成部、左下部の検索結果を表示する検索結果部と右部のキーワードの履歴などのサポート情報を掲示する履歴部の 3 つの部分によって構成される (図 1)。また、初期条件などのテキストで入力されるものは上部の入力部で行う。

クエリー作成部でノードを組み合わせるクエリーを描くと、検索結果として該当する Web ページより抽出されたキーワード群が検索結果部に提示される。結果のキーワード群はノードとして表示される。表示された検索結果のキーワード群は直接操作でき、追加するキーワードとして利用できる。

履歴部にはクエリー作成に使われたキーワードの

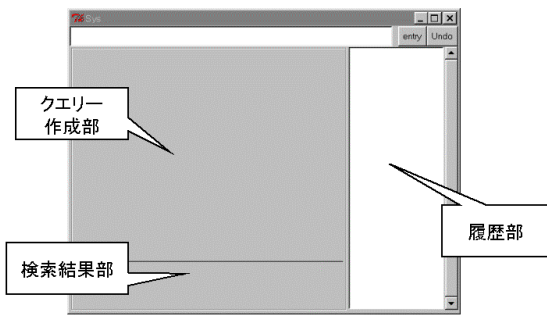


図1 システムの画面

履歴をリストとして表示する。ユーザはこのリスト内のキーワードをクリックすることでクエリーに利用できる。

利用者は表示されるキーワード群の内容によって目的のページを表示させるかどうかを最終的に決定する。

本システムにおいてユーザは以下に示す手順を踏むことによって検索を行う。

1. 検索初期条件となるキーワードを入力する(図2)。このキーワードはキーボードなどからテキストで入力するか、履歴部にある履歴から選択する。入力後、初期条件となるキーワードがノードとしてクエリー作成部に表示される。それと同時に検索が行われ、その検索結果が表示される(図3)。
2. 検索結果として画面に表示されるキーワード群の種類や個数、重要なキーワードの有無などによって検索を続けるかを判断する。
3. 検索を続ける場合、クエリーのキーワードの追加や削除を行う。追加は、検索結果として表示されたキーワード群から関係するキーワードをドラッグ&ドロップによってクエリーに組み込む(図4)。もしくは、初期条件と同様にテキストでの入力や履歴部からの選択でも可能である。削除の場合は逆にクエリーのキーワード群から不必要なキーワードをドラッグ&ドロップにより画面外へ移動させることによって削除する。
4. 目的となるようなページを示唆するようなキーワードが検索結果の中に出現した場合、検索目的が達成されたと考える。検索結果部にある検

索目的のメインとなるキーワードをクリックすることにより目的のページを表示する(図5)。

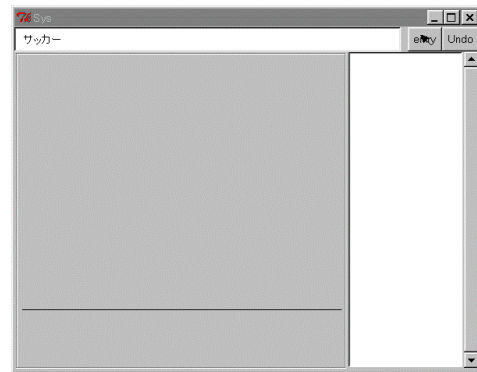


図2 検索のための操作(a)

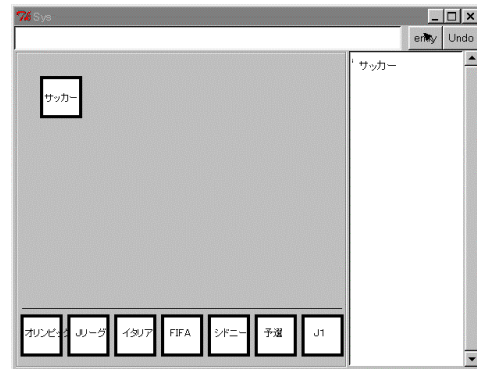


図3 検索のための操作(b)

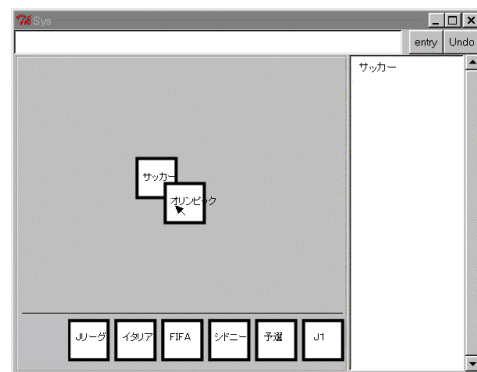


図4 検索のための操作(c)

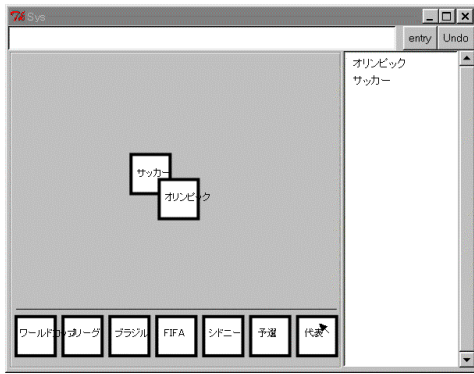


図5 検索のための操作 (d)

2.1 重ねあわせ表現による論理式

本システムでは、ユーザは基本的にノード化されたキーワードを操作することによって検索を行う。複数のノードに対し重ねあわせ表現を応用することで論理式を記述する。これにより、テキストで書かれる論理式より直観的なクエリーを作成することができる。表現としては検索において一般的に用いられる and、or を図形的に表したものをを用いる。

図6では((オリンピック and 代表) or ワールドカップ) and サッカー and 日本 の組み合わせを示している。このように or を併記で、and を重ねあわせで表現する。また、括弧を示すためにグループ化を行う。

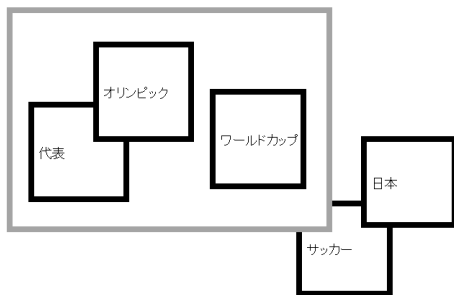


図6 重ねあわせ表現による例

2.2 過去履歴の表示

履歴部では、今までクエリーに利用してきたキーワードが過去履歴として表示される。この履歴は新しく利用されたものほど上に記述され、また、クエリー作成のためのキーワードとして利用できる。

2.3 undo機能

クエリー作成部は逆操作 (undo) をおこなう機能をもつ。これにより目的以外の検索結果に陥った場合、操作を遡ることによって検索の途中段階より再度検索を行なうことが可能になる。

右上にある undo ボタンを押すことにより、クエリーは1回前のドラッグ&ドロップ作業を行なう前の状態になる。検索結果はクエリーに基づいて逐次検索を行なっているため、基本的に同様なキーワードが表示される。また、画面右側の履歴表示は変化しない。

3 システム構成

前述の手法を用いて実際にシステムを実装している。本システムではインタフェース部と検索部に分けて実装を行った(図7)。

インタフェース部はノードの操作や図形で書き表されたクエリーの解析、検索結果であるキーワードのユーザへの提示などを行う。

検索部は検索を行うことにより検索結果であるキーワードを抽出する。すなわち、インタフェース部によってテキストに直されたクエリーを元に、URLを求める検索を行い、そのURLが示すページより関係するキーワード群を導き出しインタフェース部に返す。

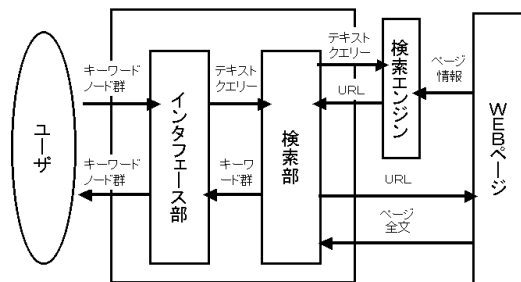


図7 システムの構成

3.1 ノード表現の変換

ノードによって表現された論理式は、そのままでは外部のサーチエンジン等にデータとして渡すことができないためサーチエンジンが利用可能な形式に変換する必要がある。

本システムでは、インタフェース部において、図

形で示されたクエリーをテキストに利用可能な中間表現に変換し、検索部に引き渡す。重ねあわせ図形を変換する際にはビジュアルプログラミングシステムの手法を用いている [1] [2]。この作業はノードの追加、削除、重ねあわせ操作があるたびに行なう。

3.2 キーワードの抽出

現在運用されているサーチエンジンのほとんどは、検索結果としてキーワードを返す機能をもたない。このため、検索部ではキーワード群を求めるために、2段階に分けて検索を行う手法を取った。1段階目ではand、or 表現で記述されたクエリーよりURLを導き出す従来の検索を行う。

2段階目では1段階目の検索結果として返ってきたURLより、該当するページを直接取得して個別に形態素解析をかけキーワードを抽出する。形態素解析には『茶笥』[3]を用いている。抽出されたキーワードは優先順位をつけ、その上位のキーワードを結果のキーワード群としてインタフェース部に渡す。

4 関連研究

検索結果をキーワードで表示して利用するものには、ArtaVista Refine [4]がある。初期条件から関連キーワードリストを作成し、利用者がそれらのキーワードを「必要」、「除外」のスイッチ選択を行うことによって、検索結果を絞りこむものである。

ノードを使った情報検索システムとして、DualNAVI [5]があげられる。DualNAVIでは特徴語グラフというもので検索対象の内容を表示する。このシステムはサーバの持つデータに対して事前に各語の全体の出現頻度を算出しておき、検索結果の語の出現頻度と比較することで、特徴語を抽出して

いる。

文章群から抽出された単語の連想関係をノードを用いて可視化するシステムとしてACCENT [6] [7]があげられる。ACCENTは検索結果である連想マップを表示する際、単語間の関係度計算やレイアウトに対してスコープを制御し結果情報の分析の目的や視点を反映させている。これにより、検索情報の分析作業の効率化を図っている。

5 まとめ

我々はノードを使った対話的キーワード検索システムを提案、実装した。このシステムでは、クエリーをノードの組み合わせによって直観的に作成できる。検索結果はURLではなくクエリーに利用可能なキーワード群で提示するため、キーワード選択の負担を軽減しWeb検索をより容易にする。

参考文献

- [1] 奥村穂高, 田中二郎: 重ねあわせを用いたビジュアルプログラミングの表記法, 日本ソフトウェア科学会第15回大会論文集, pp. 121-124 (1998).
- [2] 奥村穂高, 田中二郎: 重ねあわせを用いたシェルスクリプトプログラミング, 日本ソフトウェア科学会第16回大会論文集, pp. 61-64 (1999).
- [3] 松本裕治, 北内啓, 山下達雄, 平野義隆, 松田寛, 浅原正幸: 日本語形態素解析システム『茶笥』version 2.0 使用説明書第二版, NAIST Technical Report, NAIST-T-TR99012 (1999).
- [4] AltaVista Refine
AltaVista Company.
- [5] 西岡真吾, 丹羽芳樹, 岩山真, 高野明彦: 文献検索支援インタフェースDualNAVI, インタラクティブシステムとソフトウェアV (日本ソフトウェア科学会WISS'97), pp. 43-48 (1997).
- [6] 三末和男, 渡部勇: テキストマイニングのための連想関係の可視化技術, 情報処理学会第55回情報学基礎研究会 (1999).
- [7] 渡部勇, 三末和男: 単語の連想によるテキストマイニング, 情報処理学会第55回情報学基礎研究会 (1999).