

奥村 穂高

筑波大学大学院理工学研究科[†]

田中 二郎

筑波大学電子・情報工学系[‡]

1 はじめに

現在、Web 検索を行う手法として通常キーワードを入力し検索を行うのが普通である。しかし、テキスト入力のため絞り込むためのキーワードが選定しにくく、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。

我々はノードを利用してクエリーを生成する Web 検索システムを作成した。本論文ではそのシステムの概要について述べる。

2 ノードによるキーワード検索

ユーザは初期条件以外はすべて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。また、検索結果として Web ページの URL を返さず、検索結果の Web ページより抽出されたキーワード群を表示する。

2.1 重ねあわせ表現による論理式

本システムでは重ねあわせ表現を利用してクエリー（論理式）を記述する。

図 1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。

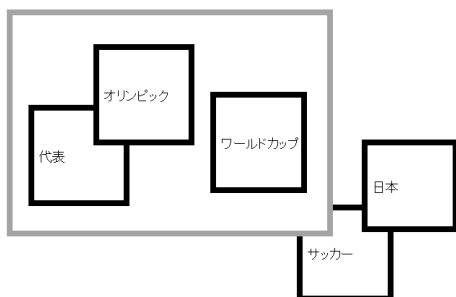


図 1: 重ねあわせ表現による例

2.2 結果のキーワード群の利用

検索初期条件となるキーワードを上部の入力欄からキーボードやカット&ペーストを利用して入力することで、中央のクエリー作成部にノードを生成する。それらのノードをドラッグ&ドロップによって組み合わせ、図 1のような表現を作成することで論理式を表現する。ノードの追加や削除、重ねあわせの発生などクエリーの図に変化がおこれば、そのたびに検索を開始する。

検索結果は Web ページより抽出されたキーワード群として画面下部に表示される（図 2下側）。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。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し、該当する URL リストを表示させる。

過去にクエリーを生成するために利用したキーワードは画面右上部に提示される（図 2右側）。また、過去の検索結果において該当したキーワードであり関連が深い現在のキーワード群に含まれていないものが画面右下部に提示される。これらのキーワードは検索結果と同様に検索初期条件やノードの生成に利用できる。

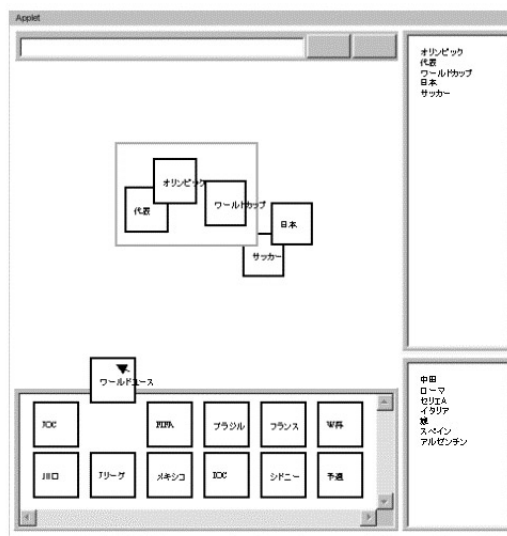


図 2: システムの実行情例

*Web Keywords Information Retrieval Using Overlapping Nodes

[†]Okumura Hotaka : Master's program in Science and Engineering, University of Tsukuba
e-mail:textex@softlab.is.tsukuba.ac.jp

[‡]Jiro Tanaka : Institute of Information Sciences and Electronics, University of Tsukuba

2.3 検索の個別カスタマイズ

クエリーとして入力したキーワード群、検索結果で求められたキーワード群についてシステムはパーソナルデータとして記録しておく。それらのキーワードについて利用頻度や検索結果内での出現頻度をしらべ、キーワード間の関連を構築する。

キーワード間の関係についてのデータは、画面右下部の関連キーワード群を生成する際に利用する。また、画面右上部にある以前利用したキーワード内で外部の検索エンジンから返された URL の中には無かったものを反転色にしたり、逆に中央部の現在クエリーに含まれているキーワードノードで効果が薄いものを色を変えて表示する、などにも利用する。

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た。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3)。

インタフェース部で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提示などを行う。また、パーソナライズなどの作業もインタフェース部が行う。

検索部ではテキストに直されたクエリーを元に検索を行い、関係するキーワード群を返す。

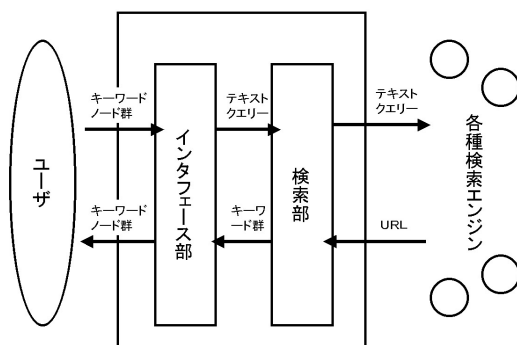


図3: システムの構成

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検索エンジン等にデータとして渡すことができないため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]。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 検索部

本システムのサーチエンジンとしては、既存のサーチエンジンを利用するメタサーチエンジンとした。既存のサーチエンジンでは検索結果の Web ページが持つキーワードは提供されないため、2段階の検索を行うことによりキーワードを抽出する仕組みになっている。

まず、中間表現の形でインタフェース部より渡されたクエリーを各種検索エンジンに渡せる形に再変換し問い合わせをおこなう。今回は Verno[2]、altavista[3]の2つの外部の検索エンジンを利用した。

検索結果として帰ってきた URL より、該当するページを直接呼び出し全文検索を行うことでキーワードを抽出する。抽出されたキーワードは複数ページへの出現などの頻度により優先順位をつけ上位キーワードをインタフェース部に渡す。

直接全文検索を行う際、キーワードや更新日時などのデータを記録し、次の検索に利用することで検索の負担を軽減する。

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。しかし、それぞれの指定がどう効いてくるかがわかりづらいため目的通りに絞り込むのが困難である。

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。このため、かなりサーバ依存のところがあり、他のサーチエンジンのデータを利用するメタサーチエンジンとするには不向きな機構である。

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] Verno
<http://verno.ueda.info.waseda.ac.jp/>
- [3] altavista
<http://www.altavista.com/>
- [4] altavista refine
http://altavista.digital.com/av/content/help_refine.htm
- [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).