

## ユーザビリティ向上を目指した検索結果の視覚化

## Visualization of search result for improving usability

垣崎 正宏<sup>†</sup>      上原子 正利<sup>†</sup>      小柳 滋<sup>†</sup>  
Masahiro Kakizaki   Masatoshi Kamiharako   Shigeru Oyanagi

## 1. はじめに

現在、膨大な情報空間から情報を見つけ出すために、検索エンジンが利用されている。代表的なものは、Google<sup>1</sup>やYahoo!<sup>2</sup>があげられる。従来の検索エンジンは一般的に欲しい情報と関連が深いと思われる単語をqueryとして入力するが、ユーザによっては必要なqueryがわからない場合がある。また1次元リストだけを結果としてユーザに返しているため、ユーザが真に求めたい情報がランクの上位にあるとは限らない。つまり、query次第で何回も再検索を行わなければならない場合がある。

現在、自分や他のユーザの過去の履歴情報に基づいて検索支援を行う研究がされているが、過去の履歴とは違った領域に関する検索を行うときや、言語の多義性により同じqueryでもユーザの欲しい情報が違った場合などではうまくいかないことが考えられる。そのため、query候補の支援と検索結果の閲覧性を向上させ、ユーザが真に求めるページに辿り着きやすくするよう、ユーザビリティの向上を果たす必要があると考える。

本研究では、検索結果ページを類似度からクラスタリングし、そのクラスタは類似度を用いて視覚化する。視覚化手法は類似度が近いものは近く、遠くのもの遠くに配置するものとする。これを行うことによってquery候補の支援、また目的のページまでのたどり着きやすさを実現でき、ユーザビリティの向上につながると考える。

## 2. 関連研究

検索は基本的にはユーザが欲しい情報と関連がある単語をqueryとして入力する形式である。しかし膨大な情報が存在するため、ヒットしたページを効率よく閲覧できるように提示する必要がある。ページを重要であると考えられる順にソートするのが一般的であるが、他にもユーザの入力したqueryにそって、検索結果を共通トピックごとにクラスタリングを行うものや入力したqueryと関連の深いキーワードを提示するものがある。また、ユーザの閲覧履歴により最適なキーワードを推薦するものもある。

こうした検索結果のページを提示するために、主に次の3つの解析がよく用いられる。

1. リンク構造
2. ページ内容
3. ユーザの閲覧

本研究では2に基づき、さらに視覚化手法を用いた検索システムを紹介する。

## 3. 視覚化を用いた検索エンジン

● Vivisimo<sup>3</sup>

Vivisimo Clustering Engine<sup>TM</sup>と呼ばれるクラスタリングエンジンを持った、メタ検索エンジン<sup>4</sup>である。検索結果のページに対し、階層クラスタリングを行うことで、意味あるグループに自動分類し、結果をツリー型メニューとして表示する。クラスタリングには、問い合わせを行った検索エンジンの出力結果から抽出したタイトル、サマリ、URLの情報を利用している。

● Grokker<sup>5</sup>

Java アプレットによる検索結果のクラスタリング表示を行うメタ検索エンジンであり、デスクトップ検索も可能となっている。リスト表示とマップ表示の2種類の表示手法を採っている。リスト表示ではVivisimoのように階層クラスタをツリー型に表示し、マップ表示では検索結果を円形のカテゴリに分類し、ページ数に応じた大きさで表示している。又、1つのカテゴリを選択すると、それに属する新たなカテゴリが表示される。

## 4. 視覚化検索エンジンの提案

検索エンジンを利用するにあたって、最もユーザの負担の大きいことはqueryの選出である。このquery次第でページに行き着くまでの効率が左右される。結果表示からさらに絞り込めるようquery候補の提示ができれば、負担は軽減されると考えられる。

Query候補の提示の方法として視覚化がある。検索結果を視覚化することにより、2次的に把握することが容易になる。しかし既存の視覚化を用いた検索エンジンではただ視覚化したただけなので、提示はできても視覚的效果は見込めない。そこで検索結果のクラスタリングを用いて、類似度によって視覚化すれば、query候補を表示しつつ、視覚化による配置からqueryの関連性を表すことができるような検索エンジンが実現できると考える。さらにページ間にも関連性を用いた視覚化を行うことで、閲覧性も向上させることができると考えられる。

## 5. システム概要

今回はYahooAPI<sup>6</sup>を利用して検索エンジンを作成する。検索結果の表示としては、リスト表示とマップ表示とする。リスト表示は、ページのランク順に類似度を付加したものをを用いる。マップ表示は、クラスタまたはページを類似度を用いて視覚化したものを表示する。

<sup>3</sup> <http://vivisimo.com/>

<sup>4</sup> 入力された索引語を複数の検索エンジンに送信し、それぞれの結果をまとめて一覧にして表示する検索サービス

<sup>5</sup> <http://www.grokker.com/>

<sup>6</sup> <http://developer.yahoo.co.jp/>

<sup>1</sup> <http://www.google.co.jp>

<sup>2</sup> <http://www.yahoo.co.jp>

### 5.1 システムの構造

検索エンジンはユーザが入力した query を含む文書をデータから高速に見つけ出す必要がある。まず、検索結果のページデータを収集し、収集されたデータに対してテキスト抽出を行い、形態素解析を行う。ここから本システムは、形態素解析されたデータをもとに特徴語を抽出し、その特徴語を用いてクラスタリングを行う。そのクラスタを視覚化してユーザに結果として表示するものとする。この一連の構造を図1に示す。



図1：検索結果表示の流れ

検索結果のページに対して、クラスタリングを行うための前処理として、特徴語・文書行列の作成を行う。この行列は特徴語集合と文書集合の関係を表した行列である。又、この処理で得られる特徴語集合は最終的に生成されるクラスタのラベル語候補となるため、適切な処理を行う必要がある。そのため検索結果のページ集合  $P = \{p_1, p_2, \dots, p_n\}$  ( $p_i$  は検索結果のページ) に対し、形態素解析を行うことで特徴語候補の抽出を行う。次いで、これら特徴語を用いて、ページ間の類似度を算出する。これらを以下に詳しく説明する。

### 5.2 形態素解析

本研究では Yahoo! API の形態素解析を用いる。これで解析された名詞を特徴語として定義する。

### 5.3 類似度抽出

類似度算出方法としてベクトル空間法を用いる[2,3]。行列として表現するため、文書を行、特徴語を列とすると、行列要素の値は対応する特徴語の出現回数となる。この行列に対して TF・IDF 法による単語の重要度を表す重み付けを行う。この行列に対してノルムによる正規化を行う。その行列を用いて内積を求めることによって、ページ間の類似度行列を作成する。

### 5.4 クラスタリング

クラスタリング手法は、階層クラスタリングを用い、最長距離法によって行う。距離は類似度を用いる。最長距離法とは2つのクラスタ間の距離として、各クラスタに属する要素間の距離のうち、もっとも遠いものをクラスタ間距離とするものである。ラベルはクラスタ間要素の中でもっとも特徴のあるものをラベルとして付けていく。

### 5.5 視覚化

視覚化するにあたってマップ側の表示アルゴリズムは、以前提案した、グラフィアウト手法[1]を用いるものとする。グラフィアウト手法はばねモデルを用い、類似度を利用して、近いページ同士の配置は近くに、遠いもの同士は遠くに配置するものである。この視覚化によって類似度の視点から視覚的効果が見込めるため、ページの探し易さや query 候補自体の関係をあらわす事ができる。リスト側の表示アルゴリズムは、ランキングを用いるものとする。ランキングアルゴリズムは、類似度を付加することで順位を変動させる。

下図に検索結果表示のイメージ画面を示す。このように、各 query の検索結果に対して階層的に視覚化する。さらに、クラスタをクリックすることで、クラスタラベルを query に追加し再検索を行う。これによってユーザビリティの向上を目指す。

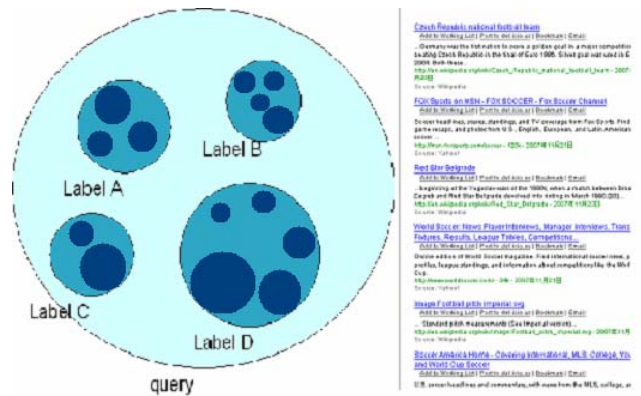


図2：検索結果表示のイメージ

## 6. 評価方法

ユーザビリティの評価においては、実際にシステムをユーザに利用してもらって、求めるページに到達するまでの操作回数を評価する。あと、処理速度の面から、紹介した検索エンジンとの比較を行いたいと考えている。

## 7. おわりに

提案した個々の手法を用いてシステムを実装することで、現在の検索エンジンの問題が解決でき、ユーザ自身の負担が軽減できてユーザビリティの向上につながると考えられる。検索時間の短縮が今後の課題である。

## 参考文献

- [1] 垣崎： “ばねモデルを用いた検索結果のグラフィアウト手法” FIT2007
- [2] 上原子正利. “関連要素決定問題の行列表現とその解法” 博士論文, 立命館大学, 9月 2006.
- [3] 徳永健伸： “情報検索と言語処理” 東京大学出版会