

ユーザの嗜好に基づく検索支援システム

A study on retrieval support systems based on user's preference

高嶋 活輝†
Katsuki Takashima間瀬 哲也†
Tetsuya Mase中山 泰一†
Yasuichi Nakayama

1. まえがき

現在インターネットの普及と共に、その利用者数が増え続けている。また、それに伴って情報量が増加し自分の最も知りたい情報を知るのが難しくなっている。

そこで検索システムが利用されるが、既存の検索システムではユーザが指定したキーワードを含んでいないもの実際には、必ずしも求める情報が得られるとは限らない。そのため大量にある Web サイトの中から自分の最も知りたい情報を得ることはかなり困難となっている。また、ユーザが検索結果を5ページ、10ページと辿って行くことは少く、再度検索を行うことがほとんどである。

そこで本研究ではユーザの嗜好を学習させ、ユーザの目的となる Web サイトを検索結果の上位に表示させる。それによりユーザの検索を支援し、求める Web サイトへのアクセスをすばやく行えるような検索システムの設計及び実装、評価を行う。

2. 関連研究

ユーザの嗜好を反映させ、検索を行うシステムには Google Personalized Search[1]がある。Google Personalized Searchは、ユーザの嗜好をあらかじめ登録しておくそのユーザに合った Web サイトを自動的に検索結果の上位へ表示するシステムである。また、その嗜好の反映度合いはスライダーによって調節できるようになっている。ただし嗜好情報の登録は手動でユーザ自らが行わなくてはならず手間がかかる。

本システムではこの手間を省くために嗜好情報の登録の過程をユーザの Web サイトの履歴より自動で行うことが可能である。

3. 設計

最適な検索を行うためにはユーザが検索の手法を学んでいくことと検索システム自体の改善が考えられる。本システムの設計方針として後者の検索システムの改善を目的とする。また、個人の嗜好を扱うということもあり、個人情報保護の観点からクライアント側で処理を行うよう設計した。ユーザの嗜好はユーザがたどった Web サイトの履歴より学習させることとした。嗜好はシステム側で自動的に学習し、ユーザの手を煩わせない仕様とした。これらを踏まえ本システムの設計を行った。その動作手順は以下のようなになる。

手順1では、実際に検索を行い、得られたオブジェクトより嗜好となるカテゴリ情報を抽出する。

手順2では、すでにユーザの嗜好を学習していた場合に検索で得られた情報とユーザの嗜好を比較する。

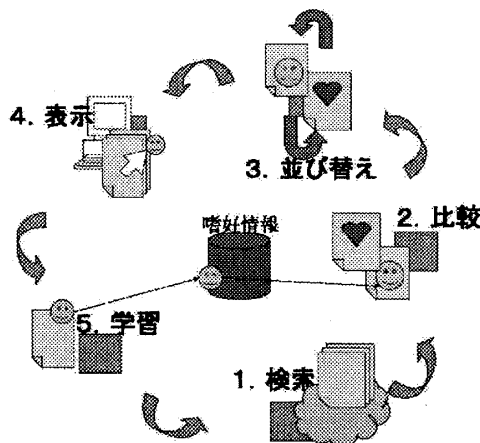


図1: 本システムの設計

次に手順3において比較した情報を元に検索結果の順序の並び換えを行う。

手順4では上で並び換えを行った通り画面上に検索結果として表示する。

手順5では、表示された検索結果をユーザが選択した場合に、選択した Web サイトのカテゴリ情報から嗜好を作成、学習させる。

本システムは以上のプロセスをくり返すことにより、ユーザの嗜好を学習し検索結果として嗜好を反映させる。この様子を示したものが図1である。

4. 実装

本システムの実装は、検索部分、並び換え部分、嗜好情報作成部分の手順に分けて行った。

検索部分の手順は GoogleWebAPI [2]によってその作業をすべて行う。また、検索によって得られたオブジェクトからその Web サイトの属するカテゴリの情報などを抽出する。

次に上で得た情報とユーザ自身の嗜好情報との比較を行い、一致したカテゴリの深さ及び嗜好の得点によって並び換えを行い検索結果として出力する。この結果一致したカテゴリの深いページがより検索結果の上位に移動することとした。

また上で得た結果を画面上に表示する。表示は並び換えられた順にタイトル、概要、URLを表示する。URL部分はユーザが選択可能なハイパーリンクとなっている。

最後に検索結果からユーザの選択した Web サイトの履歴から嗜好情報を作成する。このときの嗜好情報は Google ディレクトリサービスにおける Web サイトの分類 [3]を用いた。選択された Web サイトがすでに嗜好に

† 電気通信大学 情報工学科
Department of Computer Science, University of Electro-Communications

含まれる場合、嗜好としての得点を増加させる。並び換えはカテゴリの深さを優先して行い、得点は、2つ以上の Web サイトが同ランクにきた場合に並び換えの指標とした。

以上のプロセスを繰り返し行い、嗜好情報を作成した。またその結果を反映させ、検索支援を図る。本システムのスクリーンショットを図2に示す。

5. 評価

本システムの評価は、実際にシステムを動作させ、動作例よりその有用性を挙げることにより行った。以下にその例を示す。

今回、システムの評価に用いたのは次の3つの嗜好情報を含んだものである。

- Sports/Soccer
- Science/Astronomy
- Home/Cooking

まず、検索語「football」を入力し検索を行った結果を示す。嗜好を反映させない場合 NFL (アメリカンフットボール) に関する Web サイトが上位を占めた。嗜好を反映させた場合は、Sports/Soccer カテゴリを反映し、並び換えの際にサッカーカテゴリに属する Web サイトが上位に現れた。

次に、検索語「ATM」を入力し検索を行った結果を示す。嗜好を反映させない場合「Automatic Teller Machine (現金自動引き出し機)」としての意味の ATM がヒットし金融機関に関する Web サイトが上位を占めた。検索に嗜好を反映させると Science/Astronomy カテゴリを反映し、「Amateur Telescope Maker (アマチュア望遠鏡製作者)」としての意味の ATM がヒットし、天文に関するページが上位に現れた。

最後に、検索語「basil」を入力し検索を行った結果を示す。嗜好を繁栄させない場合バジルの植物そのものに関する Web サイトが上位を占めた。検索に嗜好を反映させると Home/Cooking カテゴリを反映し、「バジルを利用した調理方法」に関する Web サイトが上位に現れた。また、実際に検索した際の並び替え前と後の様子を比較し以下に示す。

並び替え前

1. <http://www.botanical.com/botanical/mgmh/b/basbus17.html>
2. <http://ohioline.osu.edu/hyg-fact/1000/1644.html>
3. <http://www.newadvent.org/cathen/02330b.htm>

並び替え後

1. <http://recipecircus.com/recipes/dianne/SAUCES/BasilLimeMayonnaise.html>
- 並び替え前の状態で 33 件目

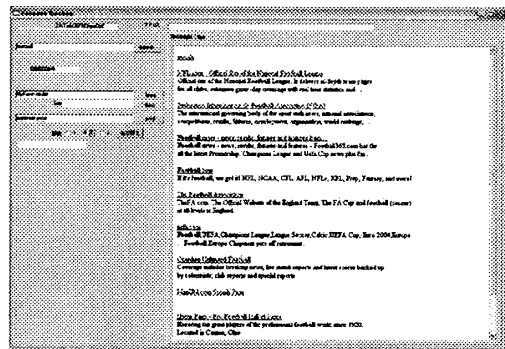


図2: 本システムのスクリーンショット

2. http://southernfood.about.com/food/southernfood/library/cprec/bl174_9.htm
同 106 件目

3. <http://southernfood.about.com/od/seasoningrecipes/r/bl131011g.htm>

並び替え前の1件目はバジルの植物に関する文書、2件目はバジルの育て方、3件目は大バジリウスという人物に関する文書であった。しかし、並び替え後は全てバジルを使用した料理に関する文書であった。また、Googleによる検索において33件目、106件目といった、いずれも通常の検索ではたどり着かないような Web サイトを表示する事が出来ている。このように、本システムは検索結果にユーザの嗜好を反映して表示させることができる。

6. 結び

本研究ではユーザの嗜好に基づく検索支援システムの設計及び実装を行った。本システムは Google Personalized Search と比べユーザの嗜好の登録の手間を省いたものとなった。検索結果としても、学習したユーザの嗜好を反映したものを表示することができ、目的となるシステムを作成することができた。また、個人情報の保護に関しても考慮したシステムを作成することができた。

また、今後はユーザの視点からの評価をしていきたい。今後の課題としてブラウザとしての機能を改善すること。また、嗜好の適応する時と場合によって検索結果を考慮することが挙げられる。

参考文献

- [1] Google Personalized Search.
<http://labs.google.com/personalized>
- [2] GoogleWebAPI.
<http://www.google.com/apis/index.html>
- [3] Open Directory Project.
<http://dmoz.org/World/Japanese/about.html>