

柔軟な検索結果再ランキングインタフェースの開発 —あらゆる検索結果を自由に並びかえる—

1. 背景

近年、多くの人々が日常的に検索エンジンを利用しており、検索エンジンは人々の情報源として欠かせないツールとなっている。しかし、現在の検索エンジンは、ユーザの検索意図にあった結果を提示することは困難である。ユーザの求める検索結果はユーザの興味や嗜好によって異なり、同じユーザが同じクエリを入力したとしても、その時の検索タスクによって求める検索結果は多岐にわたる。このような、多様な検索意図を満たすためには、ユーザは検索意図を上手く検索エンジンに伝え、精度の良い検索結果を得る必要がある。しかし、現状のシステムでは(1)検索エンジンとユーザ間のインタラクションの少なさ、(2)ユーザにとって、欲しい情報をクエリとして言語化することの難しさ、の大きな2つの問題があり、ユーザが検索エンジンに上手く欲求を伝え、求める情報を効率よく取得することは難しい。

2. 目的

本プロジェクトの目的は、ユーザが自由に自らの意図で検索結果を並べ替えることが可能なシステムを開発することである。シンプルなインタラクションを通して検索結果を再ランキングすることで、ユーザは複雑なクエリを作成することなく自由に検索結果を閲覧することができる。

3. 開発の内容

3.1 システム概要

開発したシステムはタブブラウザの1つである Firefox の拡張機能として実装されており、通常の Web ブラウジングの枠組みの中で動作する。開発システムの概要図を図1に示す。

開発システムではまず、ユーザは Google や Yahoo !、Amazon といった通常の検索エンジンにアクセスした後、クエリを入力し検索結果ページを閲覧する。ブラウザが検索結果ページを取得すると、システムは、検索結果ページの解析を試みる。検索結果ページの解析には、システム内に蓄えられた HTML パーサーのデータベースから、動作可能な HTML パーサーを調査し、該当する HTML パーサーが存在するか確認を行う。HTML パーサーが見つかったら、システムは検索結果ページを解析した後、一旦ユーザに検索結果ページをそのまま提示する。検索結果ページを受け取ったユーザは、検索結果上の気になるキーワードを選択、もしくは属性値に対してカーソルを近づけるというインタラクションを検索結果ページ上で行う。システムはユーザのインタラクションに合わせて検索結果の再ランキングを行い、再ランキングされた検索結果ページをユーザに提示する。ユーザは検索結果に満足するまで何度でも検索結果を再ランキングすることができる。

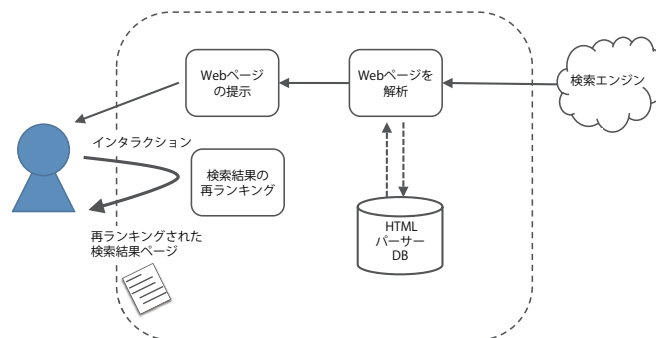


図1: システムの概要

3. 2. 検索結果の再ランキングインタフェース

ユーザは検索結果中のキーワードに対して削除操作を行うことにより「検索結果のこの部分がいらぬ」、強調操作を行うことにより「このような検索結果をもっと上位に表示して欲しい」といった意図をシステムに伝えることができる。

図 2 に強調操作を用いた検索結果の再ランキングの様子を示す。たとえば、ユーザが「豚肉とピーマンを使ったレシピが欲しい」と考え、Web 検索エンジンにアクセスし、「豚肉 ピーマン」というクエリで検索結果を取得したとする。ユーザは検索結果を閲覧してゆき、検索結果中に出現した「チンジャオロース」というキーワードに興味を惹かれ、「じゃあチンジャオロースを使った料理をもっと見てみたい」と考えたとする。その場合、ユーザは検索結果中の「チンジャオロース」を選択する。すると、システムは強調・削除というボタンをユーザに提示する(図 2 左部)。ユーザが強調を押すと、システムはユーザが現在閲覧している検索結果集合の中から、「チンジャオロース」を含む検索結果を上位に再ランキングし、ユーザに提示する(図2右部)。このように、強調操作を用いることでチンジャオロースに関する検索結果を簡単に集めることができる。この機能を用いることで、ユーザは自らの興味を簡単にシステムに伝達することができる。また、強調操作の代わりに削除操作を用いることで選択したキーワードを含む検索結果を下位に再ランキングし、上位の検索結果から除去することができる。

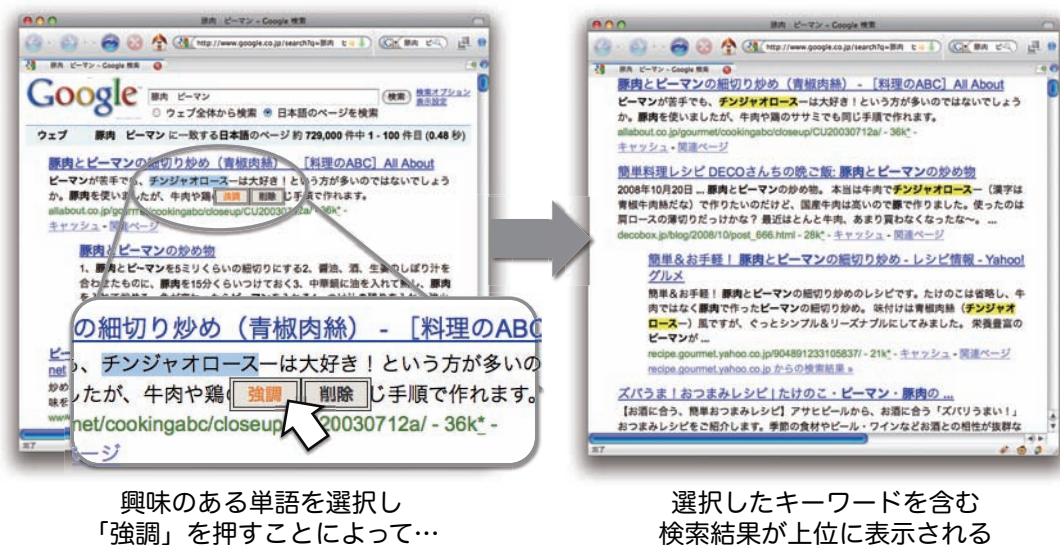


図2: 強調操作による検索結果の再ランキング例

また、キーワードベースでの再ランキングだけでなく、本システムでは数値属性(商品の価格、論文の発表年度など)に基づいて検索結果を再ランキングすることも可能である。図 3 に属性値による再ランキングの例を示す。ユーザは「Takehiro Yamamoto」に関する論文の検索結果を閲覧しているとする。「良く引用されている論文を見たい」と考えたユーザは、検索結果の「引用元」と表示されている箇所にもマウスカーソルを運ぶ(図 3 左部)。するとシステムは昇順・降順という 2 つのボタンを提示する。ユーザが降順を押すと、システムは検索結果を引用元の数で並びかえる。結果として、ユーザは論文を被引用数の多い順に論文を並べることができる。

このように、属性値に関するインタラクションにより、キーワード単位では表現することの難しい再ランキングも可能になる。



並びかえたい属性に
マウスカーソルを合わせることで

指定した属性に応じて検索結果を
昇順・降順に並べ替えることができる

図3: 昇順・降順に基づく属性値に応じた検索結果の再ランキング例

3.3 ユーザ主導の HTML パーサー作成機能

前節において、Web 検索の検索結果や、論文検索の検索結果を再ランキングする例を示した。しかし、ユーザがそうした検索結果を再ランキングするためには、システムは検索エンジンの出力する検索結果ページを解析することが必要となる。つまり、検索結果ページのどの部分が検索結果で、検索結果はどのような情報を持っているのか、をあらかじめ知っておく必要がある。ところが、検索結果ページがどのような HTML 構造になっており、検索結果中にどのような属性値が出現するかは完全に検索エンジン側に依存する情報である。そのため、ユーザは検索エンジンごとに、検索結果ページを解析するための情報 (HTML パーサー) を用意する必要がある。

しかし、一般に HTML パーサーを作成するためにはプログラミングに関する非常に高度な知識が必要とされる。HTML パーサーの数がそのままシステムの適用範囲に直結する本システムにおいては、HTML パーサーを作成する敷居が高いとそれだけシステムの適用範囲を狭めてしまうことになる。

そこで、本プロジェクトではプログラミングを必要とせず、インタラクティブに簡単に HTML パーサーを生成できる仕組みを開発した。また、ユーザが作成したパーサーを共有できるシステムについても開発を行った。この仕組みにより、ユーザ主導で検索エンジンの出力する情報を解析する HTML パーサーを作成でき、さまざまな検索エンジン上で再ランキングを行うことができる。

図4に実際に HTML パーサーを作成する様子を示す。HTML パーサーを作成するには、システムのメニューから HTML パーサーの作成を選択する。するとシステムは検索結果ページの下部に、検索結果を作成するためのインターフェースを表示する。この時、表1で示したパーサー名、動作 URL はシステムが自動的に値を挿入する。ユーザはまず、検索結果ページの中から、検索結果そのものをクリックする。システムは、ユーザのクリックした場所を検知し、クリックした箇所の DOM ツリーを表示する (図4左部)。その後、ユーザは、システムが表示した DOM ツリー上でマウスカーソルを動かす。すると、システムはマウスカーソルが現在どの要素上にあるのかを検知し、その要素で認識できる検索結果の単位を表示する。たとえば、図4左部の例ではシステムは検索結果の単位を正しく認識できていないことが分かる。そこで、ユーザはマウスを別の要素上へ動かす。すると、図4右部のように、システムは検索結果の単位を正しく認識できていることが分かる。この機能により、ユーザはシステムが正しく検索結果を認識できているかどうかを直感的に判断することができる。



図4: HTML パーサーの作成例

4. 従来の技術(または機能)との相違

本システムはユーザからの明示的なフィードバックを受け取り、検索結果の再ランキングを行うシステムである。このようなシステムは適合フィードバックシステムと呼ばれ、これまで多くの研究がなされてきた。しかし、どのようなインタフェースやインタラクションを通してフィードバックを与えるかに関してはこれまであまり注目されておらず、複雑な操作をユーザに要求するフィードバックシステムが多かった。また、多くのシステムが特殊な環境上でのみ実行され、実際の Web ブラウザ上で実行可能なシステムは多くない。

また、本プロジェクトでは、ユーザ主導の HTML パーサー作成、共有機能についても開発を行った。Web ページからコンテンツをパースする類似の仕組みはいくつか存在するが、検索結果の単位や、検索結果がどのような属性を持っているのかという部分にまで踏み込んで HTML パーサーを作成・共有できるシステムは、開発者の知る限り存在しない。

5. 期待される効果

本プロジェクトで開発したシステムにより、ユーザ主導で HTML パーサーを作成し、ユーザ側で自由に検索結果を再ランキングすることができる。つまり、ユーザは検索エンジン側のランキングに左右されず、ユーザ側の自由な観点から検索結果を閲覧することが可能となる。また、本システムで作成される HTML パーサーは、ユーザ側が検索エンジンのデータを自由に加工するための重要な情報源となる可能性がある。

6. 普及(または活用)の見通し

現在、Firefox の拡張機能として公開・開発を行っている。今後は、Internet Explorer や Safari 等の他の Web ブラウザにも移植を行っていく予定である。また、ブックマークレット形式での配布も考えている。システムの配布を通して、ユーザの操作ログを収集することを考えている。ユーザの操作履歴から、ユーザが情報検索時にどのような箇所に注目するのか、どのようなキーワードや属性に対してインタラクションを行うのかといった検索行動を明らかにしていくことができると考えている。

7. 開発者名(所属)

山本岳洋(京都大学情報学研究科 社会情報学専攻)

(参考)開発者URL

<http://www.dl.kuis.kyoto-u.ac.jp/~tyamamot/>