

検索結果のインタラクティブな評価に基づく リランキングインタフェース

軽部 孝典† 志築 文太郎† 田中 二郎†
筑波大学コンピュータサイエンス専攻†

1. はじめに

Web 検索エンジンから得られる検索結果にはユーザの検索要求に合致したページとそうでないページが混在することが多い。この原因は、検索対象分野の知識があまりないユーザにとって検索要求を単語で表現することは難しい事、一つのページ内には検索ワードを含むが他の情報も含まれている事があると我々は考える。また、同じクエリでの検索でも求める情報はユーザによっても検索の度にも違うため、ユーザは自分が欲しい情報が検索結果のランキングの上位に集めることができたなら満足のいく検索ができると考える。この問題に対し、検索結果の一つのページを検索キーとして追加指定したり [1]、検索結果のページに含まれる単語を追加選択することによって [2]、検索結果を順位付けし直すリランキングと呼ばれる手法が提案されている。これらに対して、我々は、検索要求に適合/非適合した文章片(以下、適合文章片、非適合文章片と呼ぶ)をユーザが検索結果のページ内からインタラクティブに取捨選択することにより、動的にリランキングを行えるインタフェースを提案する。本インタフェースによって、ユーザは、リランキングのための検索要求の指定と、その結果の確認をインタラクティブに一画面内で行え、徐々に検索要求を満たしてゆくことが可能である。

2. 提案インタフェース

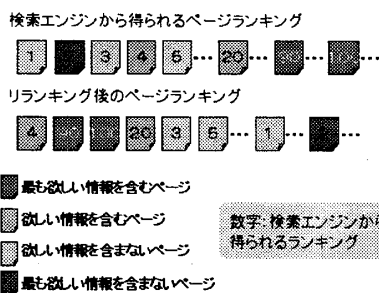


図 1: リランキングイメージ

2.1 インタラクティブに取捨選択した適合/非適合文章片によるリランキング

本インタフェースによる検索結果のリランキングのイメージを図 1 に示す。ユーザは適合文章片を検索結

A Reranking Interface based on Interactive Evaluation of Search Results

†Takanori Karube, Buntarou Shizuki and Jiro Tanaka, Department of Computer, Science, University of Tsukuba

果ページ内から選択すると、システムはその適合/非適合文章片を基にリランキングを行う。ユーザは適合文章片を追加してゆく事によって、より満足する検索要求を徐々に構築してゆく事が可能である。逆に、リランキング結果が悪くなったと感じた場合には、ユーザは先に追加した適合/非適合文章片を除けば検索要求の構成を戻す事も可能である。

2.2 全操作が一画面で行えるインタフェース

図 2 および図 3 に示す本インタフェースでは、検索結果のリランキング作業に必要な以下の操作を一画面で行うことが可能となっている。

- 検索語入力
- 検索結果リストの閲覧
- 検索結果のページの閲覧
- 適合文章片の取捨選択

ユーザが検索語入力部(図 3 参照)に検索語を与えて検索を開始すると、検索エンジンからの検索結果が検索結果リストに表示される。検索結果リストの項目を選択することによって、そのページをページ閲覧部において閲覧できる。ユーザは閲覧と同時に適合文章片の選択が可能である。表示されているページの一部をドラッグして選択し、続けて検索要求構成部へドロップすることにより適合文章片を検索要求に追加できる。この際、リランキングが行われる。なお、適合文章片は、ひとつのページの複数箇所、あるいは複数のページからいくつでも選択し追加可能である。適合文章片を検索要求から除くには、検索要求構成部にポインタを一定時間留める。すると画面構成が図 2 右のように変わり、検索要求構成部が大きく表示される。この状態の検索要求構成部にはそれまでに追加された全ての適合文章片がオブジェクトとして表示される。一度追加した適合文章片を取り除くには対応するオブジェクトを周辺領域までドラッグすれば良い。この時にもリランキングが行われる。マウスが検索要求構成部を離れると元の状態(図 2 左)に戻る。ユーザは上の (b) から (d) を高速に行えるため、検索要求を容易かつ漸次的に構築し自分の欲しいページを上位にリランクさせられる。

2.3 検索要求再構成部

検索要求再構成部では一度良い・悪いと評価したが、評価した後のランキングを見て追加する前の状態に戻りたい場合に役に立つ。いわゆる UNDO 機能で、この操作もドラッグ&ドロップによりインタラクティブに行うことができる。検索要求再構成部の概要を図 4

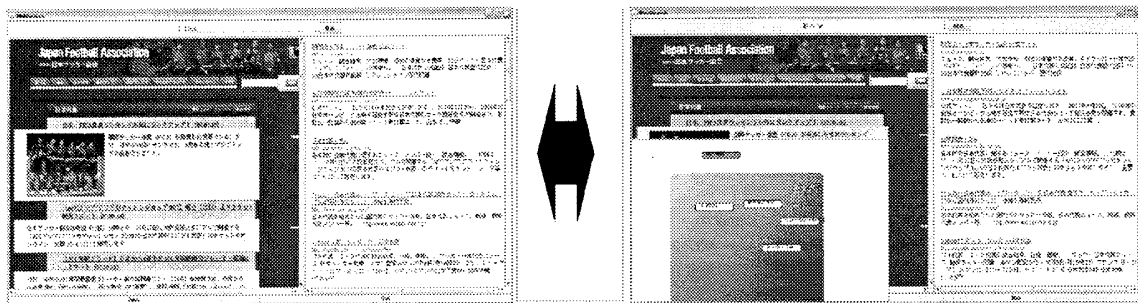


図 2: インタフェースデザイン

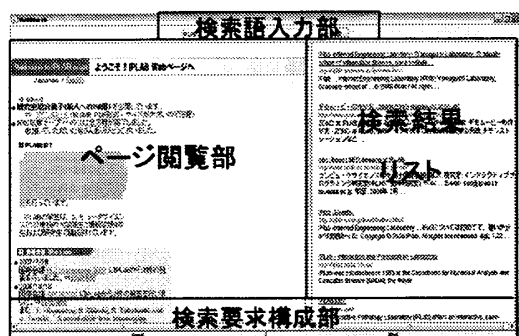


図 3: 画面構成

に示す。図 4 の内側の領域（青）が検索要求ベクトルとして計算される領域で、外側の領域（白）が検索要求ベクトルとして計算されない領域である。ドラッグ & ドロップにより出し入れ可能で、フィードバックとして計算される領域では文章オブジェクトが白で表示され、計算されない領域では赤で表示される。このように検索途中でのユーザの気持ちの変化に対応するための機能として検索要求再構成部を実装した。検索要求再構成部は、【良い】の評価と【悪い】の評価に対してそれぞれ存在し、同様の機能を提供する。

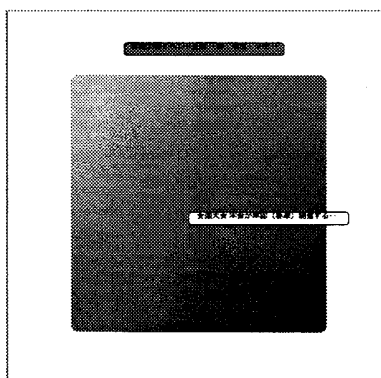


図 4: 検索要求再構成部

3. 実装

本インタフェースの実装には Java を用いた。検索エンジンとしては Yahoo! を用いている。Yahoo! の検索

API を用いて得られた検索結果から、各ページのタイトルとスニペットの名詞を抽出し、名詞それぞれに対して TF・IDF 法によって重み付けを行いその重みをベクトル空間法によって各ページの文書ベクトルを生成する。ベクトルの次元はタイトルとスニペットに含まれる全単語数 N であり、 N 次元空間に配置される。また、リランキングのために、追加された適合/非適合文章片全てから名詞を抽出し、文書ベクトルと同様にベクトルを生成する。リランキングの際には、各ページの特徴ベクトル（文書ベクトル）と検索要求ベクトルのコサイン類似度を計算することで、検索結果を並べ替える。

4. 関連研究

山本らによる研究 [3] では検索結果に対するユーザの操作を用いてリランキングを行うという点で本研究と似たアプローチを取っている。しかし一方でクエリとして追加、削除されるのはタイトルかスニペット内の単語であり、またボタンを介しリランキングを行っている。これに対し本研究は文章片を用いている点とインタフェースデザインが異なる。

5. まとめと今後の課題

検索要求を検索結果のページ内の文字列を適合/非適合文章片として選択することにより表現する手法、および検索要求を検索要求ベクトルに追加するインタラクティブな手法、インタラクティブ性を実現するための表示方法の特徴とするインタフェースを実装した。

参考文献

- [1] 平田 陽一, 松倉 健志, 田島 敬史, 田中 克己: Web 検索における意味的適合フィードバック機構, 情報処理学会研究報告, Vol.2000, No.69, pp. 137-144, 2000.
- [2] 酒井 浩之, 大竹 清敬, 増山 繁: 絞り込み語提示による一検索支援手法の提案, 言語処理学会第 7 回年次大会 発表論文集, pp. 185-188, 2001.
- [3] 山本 岳洋, 中村 聡史, 田中 克己: 編集操作を用いたウェブ検索結果の最適化, 第 17 回データ工学ワークショップ (DEWS2007) 論文集, L4-3, 2007.