

## Web 文書への主観的評価を融合しての検索結果の順位決定方式

高山毅<sup>†</sup>岩手県立大学ソフトウェア情報学部<sup>†</sup>

## 1 はじめに

従来, Web 文書の検索結果の順位は, Google の場合の PageRank[1]を代表として, 「Web 文書の被リンク数」や「Web 文書内でのタグ構造の正当性, あらわれる単語等の情報」から, 客観的かつ機械的に決定されている. しかしながら, 自身の Web 文書の順位を少しでも上げるべく, 検索エンジン最適化の取り組みが活発化しており, ユーザが真に必要なとする情報へ辿り着くことが容易とは言えない状況になっている[2].

本研究は, ユーザが真に必要なとする情報を Google を代表とする従来の検索エンジンよりも容易に得られることを目指している. 具体的なアプローチとしては Web2.0[3]の知見を生かし, 従来手法の客観的かつ機械的な順位付けへ, 各 Web 文書の主観的な順位付けを融合する.

## 2 Web 検索結果の順位決定方式

近年, Web 文書の検索技術の進歩はめざましく, 様々な順位決定アルゴリズムが考えられ, 実際の検索エンジンに用いられている[4].

**手法1:**文献[5]では, ある Web 文書のソーシャルブックマークへの登録数とその有用度には相関があると考え, SBRank という尺度を定義している. そして, SBRank と PageRank を融合して検索結果の順位を決定し, 実験によりその有効性を示している.

**手法2:**文献[4]では, 検索結果中でユーザが不要と考える単語に「削除」操作を, 必要と考える単語に「強調」操作をできるインタフェースを開発し, 検索結果をリランキングする手法を提案している. そして, 実験によりその有効性を示している.

**手法3:**著者らの研究グループでは, 検索結果をグループ分けする研究に取り組んでいる. 具体的には, 文献[6]で従来のグループ分け手法である分類やクラスタリングとは異なる, どんな検索に対しても慣れ親しんだカテゴリ名を用いてグループ分けできる, ページタイプを用いたグループ分けを提案し, 所望の情報に効率良く辿り着くことを試みている.

**手法4:**Yahoo! MyWeb2.0 (<http://myweb2.search.yahoo.com/>)の「ソーシャルサーチ」では, 気に入った検索結果をコメントやタグとともにユーザの PC に保存できる. そして, 同じ趣味嗜好を共有するメンバー間で各自が保存した Web 文書を対象に検索

を行なうことで, より信頼性の高い検索を実現しようとしている. これは, 保存した人が居た Web 文書を上位に持ってくる順位決定方式と見なすこともできる.

**各手法の問題点:**手法1, 4では「網羅性」が問題で, あらゆる検索課題に対して各手法を効果的に働かせることは容易とは言えない. また, 未登録の Web 文書は最低ランクの評価になってしまうので, 新しい Web 文書が表示されにくくなってしまいう危険性を否定できない. 手法2では, Web 文書の有用度はある単語が含まれるか否か以外にも影響を受けることである. 有用な Web 文書を下位へ, 有用でない Web 文書を上位へリランキングしてしまう危険性は否定できない. 手法3では, 各グループ内での順位付け方式は, 充分とは言えない.

## 3 Web 文書への主観的評価の融合

## 3.1 主観的評価の導入

現在, インターネットの世界では「Wikipedia」のような必ずしも客観性のみで固執せずに集合知を有効利用する Web2.0 の研究が活発化している. これは, Web の閲覧者が各自の持つ知識を持ち寄り, それらを有機的に融合することによって価値ある情報を生み出そうとするものである.

本研究では検索結果の順位決定にユーザの主観的評価を有効利用することを試みる. 図1に, 本稿で提案する「Web 文書への主観的評価を融合しての検索結果の順位決定方式」を示す. 融合の配分比は, 画面上から柔軟に設定できるようにする.

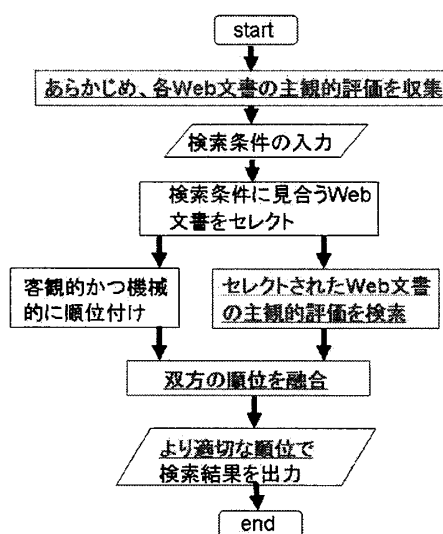


図1 Web 文書への主観的評価を融合しての検索結果の順位決定方式.

Ranking Method of Web Search Results by Incorporating Subjective Evaluation Value to Web Document

<sup>†</sup>T. Takayama(Faculty of Software and Information Science, Iwate Prefectural University)

### 3.2 主観的評価の収集

本研究ではまず、主観的評価を融合して Web 検索結果の順位を決定することがどの程度有効であるのかを、 $Q_s$  個の検索課題を設定して試験的に評価する。具体的には、検索課題を設定した上で、 $Subj$  名の被験者に各自の検索結果にあらわれる Web 文書群の上位  $TopN$  件に対して、主観的評価を  $Mul$  段階評価値の付与により依頼する。ここで  $Mul$  は奇数とし、 $\pm 0$  を「どちらとも言えない、未評価」に相当する中心として、プラス評価、マイナス評価のいずれも可能にする。その際、「(i) 特定されている検索課題に対して有用か否か」「(ii) 検索課題を問わず有用か否か」の二つの判断基準が考えられる。ここでは、Web 文書に付与した主観的評価を種々の検索課題で利用することを目指し、(ii) の判断基準で評価値を付与することにする。Web 文書の数は膨大であり、検索課題やクエリによって、あらわれる Web 文書は変化する。しかし、 $Subj$  名に同一の検索課題を与えることで、関連する Web 文書に対して主観的評価を与える被験者が極端に少なくなる可能性を緩和する。

### 3.3 主観的評価の補完

繰り返しになるが、Web 文書の数は膨大である。あらかじめ主観的評価の付与作業を行なったとしても、評価値の付与者と異なるユーザが同一の検索課題に対して検索を実施すると、評価値が付与されていない Web 文書が出現する危険性を否定できない。ここで、未評価のものの評価値を  $\pm 0$ 、すなわち評価値の中位に位置付け、最低評価に陥ることを回避する。また、将来的な方向としては、ユーザがある Web 文書を主観的に有用と評価した時に取る行動としては、

**行動 1** : ブックマークに登録する

**行動 2** : 印刷する

**行動 3** : Web 文書の部分集合をマウスでドラッグして、エディタ等にコピーする

等が考えられる。そこで主観的評価の付与方法を前述のような明示的なもの以外に、暗示的なものまで拡張する。具体的には、

**暗示付与 1** : 文献 [5] や Yahoo! MyWeb2.0 の「ソーシャルサーチ」の考え方を拡張し、個人限定まで含めてブックマーク登録されているものは、登録数に応じて評価値  $Mul$  段階中でプラス方向の評価値を付与する。

**暗示付与 2** : 印刷されたものは、印刷回数に応じて評価値  $Mul$  段階中でプラス方向の評価値を付与する。

**暗示付与 3** : マウスでドラッグの上、コピー&ペーストされたものは、その回数に応じて評価値  $Mul$  段階中でプラス方向の評価値を付与する。

なお、不当な主観的評価値向上を防ぐための対策を、それぞれ検討する。

### 3.4 システムの実装方式

SlothLib [7] を用いてシステムを開発する。SlothLib は検索エンジン開発のためのプログラミングライブラリであり、比較的低コストで所望の検索エンジンを開発できる。検索結果の順位付けは、デフォルトでは客観的かつ機械的な評価に基づいている。開発するシステムは、その上に主観的評価を融合する部分をかぶせる形で実現する。

### 4 まとめと今後の展望

本稿では、Web 文書への主観的評価を融合して検索結果の順位を決定する方式について述べた。本提案は、範囲を限定して主観的評価を利用する digg.com (<http://www.digg.com/>) や reddit.com (<http://reddit.com/>) の取組みを拡充して、より一般化を目指すものである。ユーザが真に必要な情報を得るための労力を軽減できれば、ユーザは創造性を必要とする知的な活動へ、より多くの精力を傾けることが可能になる。今後の展望は、以下の通りである：(i) 現在開発中の試作システムの完成、(ii) どの程度主観的評価を融合するのが効果的かの分析。

### 参考文献

- [1] Page, L., Brin, S., Motwani, R., and Winograd, T.: "The PageRank Citation Ranking: Bringing Order to the Web," in *Proc. of the WWW Conference*, pp. 161-172, 1998.
- [2] 山名早人: 特集 検索エンジン 2005—Web の道しるべ— 1 検索エンジンの概要、情報処理, Vol. 46, No. 9, pp. 981-987, 2005.
- [3] O'Reilly, T.: What is Web2.0: Design Patterns and Business Models for the Next Generation on Software, <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web20.html>, 2005.
- [4] Yamamoto, T., Nakamura, S., and Tanaka, K.: "Rerank-By-Example: Efficient Browsing of Web Search Results," in *Proc. of the 18th International Conference on Database and Expert Systems Applications (DEXA2007)*, pp. 801-810, 2007.
- [5] Yanbe, Y., Jatowt, A., Nakamura, S., and Tanaka, K.: "Can Social Bookmarking Enhance Search in the Web?," in *Proc. of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2007)*, 2007.
- [6] 金子大輔, 高山毅, 池田哲夫, 長内亘: "Web 文書のページタイプを用いた適応的分類と試作システムの評価", 知能と情報(日本知能情報ファジイ学会誌), Vol. 18, No. 2, pp. 319-336, 2006.
- [7] 大島裕明, 中村聡史, 田中克己: SlothLib: Web 検索研究のためのプログラミングライブラリ, 日本データベース学会 Letters (DBSJ Letters), Vol. 6, No. 1, 2007.