

# Web 文書への主観的評価を融合する Web2.0 型検索エンジンの開発

田中直樹<sup>†</sup> 高山 毅<sup>†</sup> 佐藤永欣<sup>†</sup> 村田嘉利<sup>†</sup> 生田脩二<sup>†</sup>

岩手県立大学ソフトウェア情報学部ソフトウェア情報学科<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(SEO)の取り組みが活発化しており, ユーザが真に必要な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. 著者らの研究グループでは2章で述べるように, 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文書の主観的な評価値を融合する方式を提案している[4][5]. 本稿では本方式を盛り込んだWeb2.0型検索エンジンを開発する.

## 2. Web文書への主観的評価を融合するWeb2.0型検索エンジン

図1に, 「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.

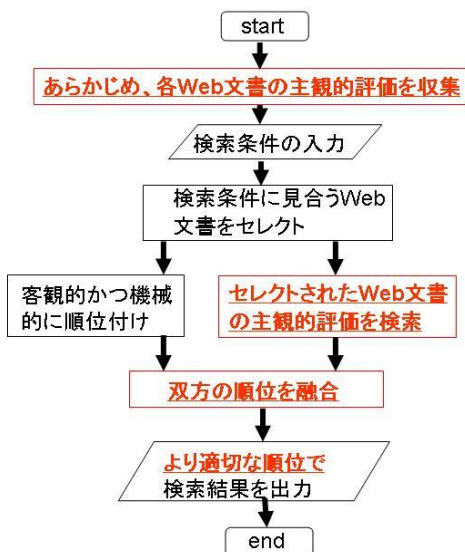


図1 Web 文書への主観的評価を融合しての検索結果の順位決定方式.

## 3. Web 文書への主観的評価を融合する Web2.0 型検索エンジンの開発

### 3.1 主観的評価値の取得

文献[4]で述べている通り, 各 Web 文書の主観的評価値  $S$  は, 後述する「現実性を高める手法」までの暫定的な方法として, 被験者実験により取得する. 被験者にあらかじめ用意したいくつかの検索タスクを課し, 検索結果としてあらわれた Web 文書に評価値の付与を依頼する. ここで主観的評価値は, 以下の二つに分けて考えることができる:

- (i) 対タスク有用性  $T$ : 課されている検索タスクに対して有用か否か
- (ii) Web 文書有用性  $W$ : 検索タスクを問わず有用か否か.

いずれも,

- +3: 非常に有用である
- +2: 有用である
- +1: 少し有用である
- ±0: どちらとも言えない
- -1: あまり有用でない
- -2: 有用でない
- -3: まったく有用でない

の七段階のいずれかで, 評価値の付与を依頼する. 本研究では主観的評価値  $S$  として, 対タスク有用性  $T$  と Web 文書有用性  $W$  を状況によって使い分ける. すなわち主観的評価値  $S$  の定義域は,  $-3 \sim +3$  である.

### 3.2 客観的評価値の取得

開発するシステムでは, 基礎となる検索エンジンとして, Yahoo![6]を使用することとする. Yahoo!検索 Web API[7]では 2010 年 11 月現在, Yahoo!で検索した結果の上位 20 件を, 「Yahoo!側で定めた客観的かつ機械的な順位  $Y\_Ord$ 」付きで取得することができる. そこでこの 20 件からその Web 文書の順位を引き,

$$\text{客観的評価値 } K = 20 - Y\_Ord \quad (1)$$

を, 客観的評価値として採用する. すなわち, 客観的評価値  $K$  の定義域は  $0 \sim 19$  である.

### 3.3 客観的評価と主観的評価の融合方法

3.1, 3.2 節で述べたように, 主観的評価値  $S$  と客観的評価値  $K$  では, 定義域が異なる. したがって双方の融合にあたり, 各評価値のままの

Development of Web2.0 Type Search Engine Incorporating Subjective Evaluation Value to Web Document  
<sup>†</sup>N. Tanaka, T. Takayama, N. Sato Y. Murata and S. Ikuta  
 (Faculty of Software and Information Science, Iwate Prefectural University)

値で和や積を取ると、一方にやや重みを置くことになってしまう。そこで本稿では、それぞれの評価値を 0.0~1.0 に正規化して、その後に融合することにする。具体的には、正規化した主観的評価値を  $S'$ 、正規化した客観的評価値を  $K'$  とするとき、

$$S' = (S - S \text{ の定義域の下限}) / S \text{ の定義域の幅} \\ = (S + 3) / 6.0 \quad (2)$$

$$K' = (K - K \text{ の定義域の下限}) / K \text{ の定義域の幅} \\ = K / 19.0 \quad (3)$$

である。

一方、主観的評価と客観的評価の各重みを、検索エンジンの閲覧者が自身の判断で自由に変われば、有用である。そこで、主観的評価の割合を  $\alpha$  ( $0 \leq \alpha < 1$ ) としたとき、融合された尺度での評価値 merge を、

$$\text{merge} = \alpha S' + (1 - \alpha) K' \quad (4)$$

で算出することを提案する。なおここで  $\alpha = 1$  を含めないのは、以下の理由による。すなわち、Yahoo! 検索 WebAPI から取得した検索結果 20 件の順位を並び替える以上、客観的な評価の割合は 0% より必ず大きくなるためである。

### 3.4 画面イメージ

図 2 は試作システムの画面イメージである。クエリを入力し、主観的評価の割合を選択して検索ボタンをクリックすると、検索結果が表示される。検索結果として返されるのは、Yahoo! 検索 WebAPI から取得された 20 件であるが、主観的評価の割合によって順位が変化する。

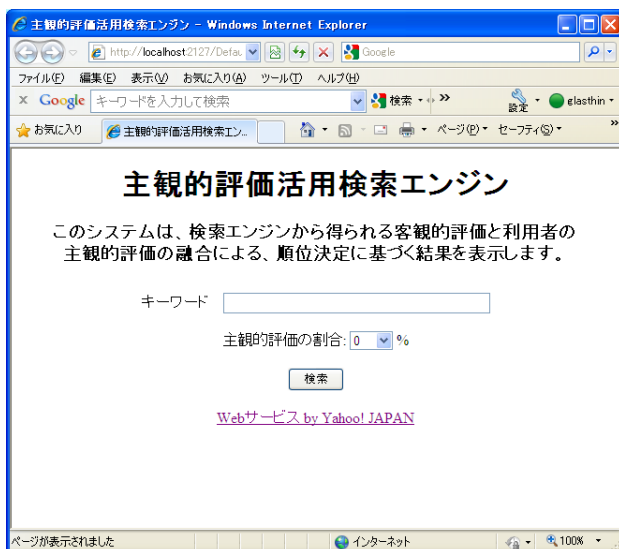


図 2: 試作システムの画面イメージ図。

## 4. 結論と今後の展望

本稿では、Web 文書の客観的かつ機械的な順位付けへ主観的評価を融合する Web2.0 型検索エン

ジンの開発について述べた。

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を完成させた上で、その有用性を評価すること、(ii) 3.1 節で述べた「現実性を高める手法」の開発。具体的には、閲覧者が Web 文書を有用と評価したときにしばしば取る行動である、「印刷」「ブックマークへの登録」「Web 文書の一部をマウスでドラッグして、エディタ等にコピー」等の行動をしたときに、その Web 文書の主観的評価を自動的に良くするインタフェースの開発。

### 謝辞

本研究の一部は、独立行政法人日本学術振興会科学研究費補助金基盤研究(C)「ポスト Google 時代に対応する Web2.0 型検索エンジンの開発」(課題番号 20500097) による。ここに謝意を表す。

### 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] 高山毅: 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, 情報処理学会第 71 回全国大会, 2C-7(2009).
- [5] 高山毅: Web 文書への評価値を融合しての検索結果の順位決定 ~同一検索タスクへのユーザ間の Web 文書の重複度調査~, 情報処理学会第 72 回全国大会, 3C-3(2010).
- [6] Yahoo! JAPAN: <http://www.yahoo.co.jp/>
- [7] Yahoo!デベロッパーネットワーク - 検索: <http://developer.yahoo.co.jp/webapi/search/>