

機関番号：21201  
 研究種目：基盤研究(C)  
 研究期間：2008～2010  
 課題番号：20500097  
 研究課題名(和文) ポストGoogle時代に対応するWeb2.0型検索エンジンの開発  
 研究課題名(英文) Development of Web 2.0 Type Search Engine for Post Google Times  
 研究代表者  
 高山 毅 (TAKAYAMA TSUYOSHI)  
 岩手県立大学・ソフトウェア情報学部・准教授  
 研究者番号：70275402

## 研究成果の概要(和文)：

従来、Web文書の検索結果の順位は、Googleの場合のPageRankを代表として、客観的かつ機械的に決定されている。しかし、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、ユーザが真に必要とする情報へ辿り着くことは容易とは言えない。本研究課題では、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、Web2.0型検索エンジンを開発した。

## 研究成果の概要(英文)：

In usual, retrieval result ranking of web documents is determined in an objective and machinery manner, typically such as 'pagerank' in the case of 'Google'. However, it becomes active to trying search engine optimization in order to obtain better ranking towards one's own Web document. As a result, it comes to be difficult for a retrieval user to obtain just a suitable information. This project has developed a Web 2.0 type search engine which incorporates subjective evaluation of Web document into objective and machinery ranking of conventional method.

## 交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	800,000	240,000	1,040,000
2009年度	500,000	150,000	650,000
2010年度	400,000	120,000	520,000
年度			
年度			
総計	1,700,000	510,000	2,210,000

研究分野：総合領域

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：情報システム、検索エンジン

### 1. 研究開始当初の背景

従来、検索結果のランキングは、Google の場合のページランク [Page98] を代表として、「Web ページの被リンク数」や「Web ページ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た [山名 05]。しかしながら、自身の Web ページのランキング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、ユーザが真に必要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た [山名 05]。

研究代表者は、この問題への一つのアプローチとして、検索結果をグループ分けする研究に取り組んでいた [金子 06]。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であった。しかしながら、この種のグループ分けのアプローチのみでは、真に必要とする情報に効率的に辿り着くには限界があった。

研究開始当初、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ならず集合知を有効利用する Web2.0 [0' Reilly05] の研究が活発化していた。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った。実際、Wikipedia に掲載されている情報は、国語の辞書として定評のある広辞苑と比較しても、質的に決して劣らない高品質な情報であることが明らかになってきていた。

一方、E-コマース他の情報推薦（レコメンデーション）で用いられる協調フィルタリング [Goldberg92] も、ユーザの主観的評価を有

効利用していた。これは、ユーザを相互の類似度に基づいてグループ分けし、類似のユーザは同一の対象に対して類似の評価を下すという仮定の下に、未評価の対象に対する評価を予測して、情報推薦を行なうものであった [土方 04]。情報推薦に関する研究も活発化して、すべてのユーザに対して同一の情報を推薦するのみならず、ユーザの特性に応じて推薦する情報を変更する取り組みも活発化していた [Riecken00]。

以上の研究動向および自身の研究成果を踏まえ、研究代表者は検索結果のランキング決定にユーザの主観的評価を有効利用する着想に至った。

### 2. 研究の目的

本研究の目的は、ユーザが真に必要とする情報を 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ようにすることである。

### 3. 研究の方法

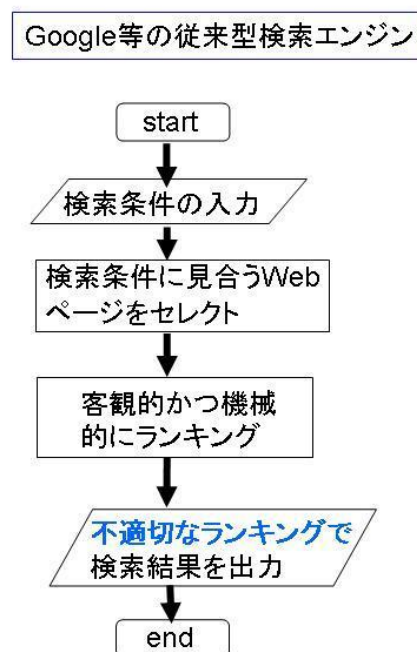


図1 従来の検索エンジン.

従来の検索エンジンが行なってきた、各 Web ページの客観的なランク付け (図1) に加え、「Web2.0」と「E-コマース等で用いられている協調フィルタリング」の知見を生かし、各 Web ページの主観的なランク付けを融合する (図2)。そして実際に、これに基づく試作システムを開発する。検索結果のランキングの決定に主観的評価を利用する点に、学術的な特色がある。

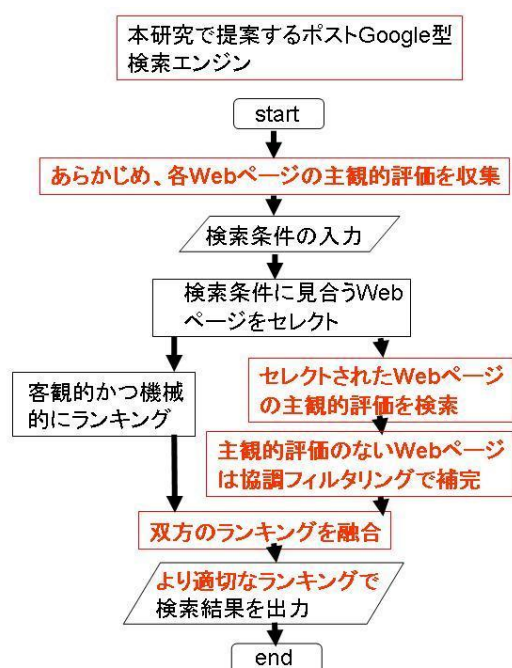


図2 本研究で提案する検索エンジン.

#### 4. 研究成果

提案手法にしたがった検索エンジンを開発した。また、主観的評価を融合すると適切な Web 文書を検索結果の上位にランキングすることが可能なことを確認できた。

開発では、基礎となる検索エンジンとして当初は Google を予定していたが、予定を変更して Yahoo! を採用した。これは、Google WebAPI によって得られる客観的評価値は、実際の検索エンジン Google で使われているも

のとは異なり、小数点以下切捨てのラフなものであり、客観的評価値に基づいて各 Web 文書の順位付けを行なおうとする時に、同順位のものが増発して不都合なためである。Yahoo! 検索 Web API 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、客観的評価値  $K=20-Y\_Ord$  を、客観的評価値として採用した。すなわち、客観的評価値  $K$  の定義域は  $0 \sim 19$  である。一方、主観的評価値  $S$  は、これまで述べてきた通り  $-3 \sim +3$  の 7 段階を採用した。すなわち、主観的評価値  $S$  の定義域は  $-3 \sim +3$  である。 $S$  と  $K$  をそれぞれ値域が  $0.0 \sim 1.0$  になるように正規化してから融合し、提案方式での検索結果の順位を定める融合評価値として採用した。提案手法により、ユーザが真に必要とする情報を得るための労力を軽減できることが確認できた。

#### 参考文献

- [Page98] Page, L., Brin, S., Motwani, R., and Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web, *Proc. of the WWW Conference*, pp.161-172, 1998.
- [山名 05] 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.
- [金子 06] 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ィ学会誌), Vol.18, No.2, pp.319-336, 2006.
- [O' Reilly05] 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.

[Goldberg92] Goldberg, D., Nichols, D., Oki, B. M. and Terry, D.: Using collaborative filtering to weave an information Tapestry, *Comm. of the ACM*, Vol.35, No.12, pp.61-70, 1992.

[土方 04] 土方嘉徳：情報推薦・情報フィルタリングのためのユーザプロファイリング技術，人工知能学会誌，Vol.19, No.3, pp.365-372, 2004.

[Riecken00] Riecken, D., *et. al*:  
“Personalized Views of Personalization,” *Comm. of the ACM*, Vol.43, No.8, pp.26-158, 2000.

## 5. 主な発表論文等

[学会発表] (計3件)

[1] 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，於立命館大学、2009年3月10日。

[2] 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，於東京大学、2010年3月10日。

[3] 田中直樹，高山毅，佐藤永欣，村田嘉利，生田脩二：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、情報処理学会第73回全国大会、2P-4、於東京工業大学、2011年3月3日。

## 6. 研究組織

(1) 研究代表者

高山 毅 (TAKAYAMA TSUYOSHI)

岩手県立大学・ソフトウェア情報学部・准教授

研究者番号：70275402