

Web 情報の多面的集約による不確かな命題の信頼性判断支援

山本 祐輔 手塚 太郎 アダム ヤトフト 田中 克己†

† 京都大学大学院情報学研究科

1 はじめに

ある知識の真偽について調べたいとき、Web 検索エンジンを用いてそれを調べることは広く行われている。この際、ユーザの多くはその知識に関するキーワードを Web 検索エンジンに入力し、関連するページの数や内容を調べたり、その知識に関して信頼のおける Web ページを探そうと試みる。しかしこのような作業はページを何度も比較しなければならず非常に労力がかかる。

そこで我々は不確かな知識の信頼性判断を支援するためのシステム「ほんと?サーチ」を開発している [1]。ほんと?サーチでは Web 上には存在する膨大な情報を効率的に集約することで、信頼性判断支援に役立つ情報を提示する。本研究では、「多くの人が認知している知識を評価する」という基本的な評価軸の基、さらに知識に対する印象、認知度の時間的変遷の観点からの Web を集約・分析した情報を提示する手法を述べる。

2 搭載機能

本章ではほんと?サーチに関する機能を説明する。ほんと?サーチではユーザはまず不確かに思う知識をフレーズの形で入力する。フレーズが入力されると、システムは入力知識と類似、対立する解候補知識を生成する。その後、それら解候補知識の信頼性を判断するための情報を Web から集約し提示する。判断情報は前章で述べた基準に基づき以下の付加情報として提示する。これらを実現するため、システムの動作フローは図 1 のようになっている。

- 単純認知度 (知識を言及している Web ページ数)
- プラス/マイナス印象の分布グラフ
- 認知度の時間変化を示すグラフ

Supporting Judgement of Trustworthiness on Knowledge by Search Results Aggregation from Sentimental and Temporal Viewpoint

Yusuke YAMAMOTO, Taro TEZUKA, Adam JATOWT and Katsumi TANAKA†

Graduate School of Informatics, Kyoto University Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan
{yamamoto, tezuka, adam, tanaka}@dl.kuis.kyoto-u.ac.jp

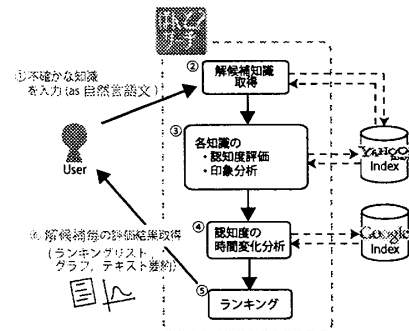


図 1: ほんと?サーチのシステムフロー

2.1 Web における知識の認知度

本システムでは入力した知識の特定の部分をユーザに指定させることで、システムはその箇所に意味的に適合するような別のキーワードを Web から取得する。取得されたキーワードと元の入力された知識から復元された知識を「解候補知識」と定義する。解候補知識を Web から抽出する具体的な手法を図 2 に記す。

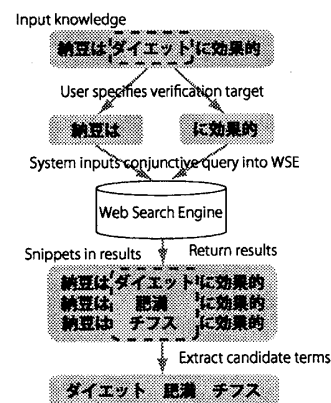


図 2: 解候補知識の抽出手順

Web にはユーザの意見、社会現象が反映されている。どの程度の数の人々が言及しているかが分かれば、知識やトピックに対する社会における認知度を近似的に求めることが可能である。そこで先の手順で得られた解候補知識をクエリとして検索エンジンに与えることで得られたページ数を知識に対する認知度とする。

2.2 知識に対する印象の集約

前章で評価した知識はフレーズをクエリとして Web 検索エンジンを用いて検索した際のヒット件数を用いているので、知識に対してプラスの印象とマイナスの印象を持つページが混在しても同等に集約してしまっている。そこで熊本らの [2] を応用して、検索エンジンの返す結果に含まれる検索結果要約 (スニペット) の印象を分析することで、Web 上である知識がプラス/マイナスの印象を持たれているかを集約する。

図 3 に「納豆はダイエットに効果的」という知識をシステムに入力し、印象分析した例を示す。ほんと?サーチは解候補として「ウォーキングがダイエットに効果的」「コーヒーがダイエットに効果的」「運動がダイエットに効果的」など計 17 個の知識を Web から抽出した。図 3 は単純認知度によるランキング結果である。この結果からは 1 位は 236 ポイントで「ウォーキング」になっており、「納豆」はスコア 51 ポイントで 8 位になっている。図からも分かるように、1 位、2 位の結果に比べ「納豆」のネガティブ印象の割合が大きい。

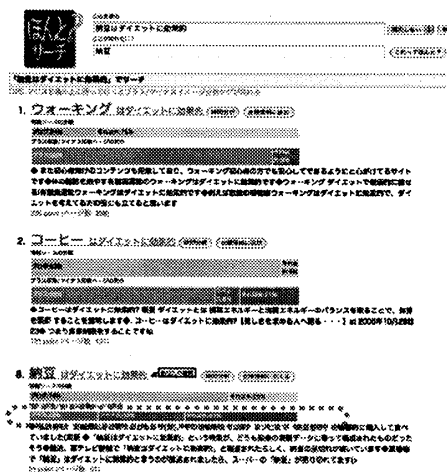


図 3: 単純認知度の評価と印象分析

2.3 認知度の時間変化

知識の信頼性は時間に強く依存することから、時間を考慮せずに Web 情報を集約し、算出された知識の認知度を信頼性判断支援に用いるのは適切ではない。そこでほんと?サーチでは集約に用いる Web ページを時期を区切って収集することで、時期毎の認知度を評価しグラフ描画する。これによりある知識は「長い間受け入れられ続けてきた知識」「最近急に認知されるようになった知識」かどうかをグラフを見ることで確認することができる。時期毎のページの収集は Google の時間指定検索オプションを使うことによって実現できる。

図 4 に「ウォーキングはダイエットに効果的」と「納

豆はダイエットに効果的」の時間的な認知度変化を比較した例を示す。この結果からウォーキングとダイエットでは認知度の変化が異なり、納豆は 2 度ブームが起こっており、「一回目のブームと 2 回目のブームで印象の変化が起きた」などの推測が可能となる。

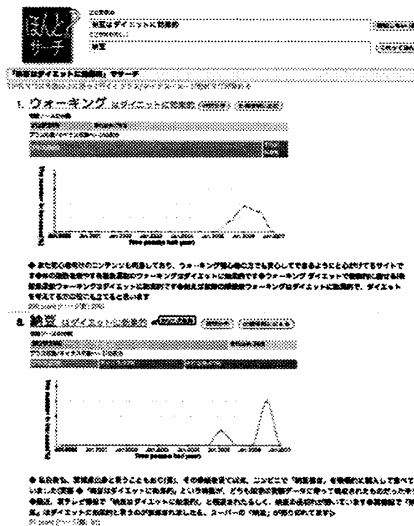


図 4: 認知度の時間的変遷のグラフ提示

3 おわりに

知識の信頼性を調べたいときにそれを判断支援するシステムは現状では存在しない。それゆえ我々は検索エンジンの結果を集約することで、プラス/マイナス印象、時間的な観点から知識の信頼性を判断支援する手法を提案した。

4 謝辞

本研究は一部、グローバル COE 拠点形成プログラム「知識循環社会のための情報学教育研究拠点」、文科省研究委託事業「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」、科研費：計画研究「情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究」(課題番号 18049041)、若手研究 (B)「ウェブ活用のための情報統合による信頼性判断支援」(課題番号: 18700086)、若手研究 (B)「情報検索とウェブアーカイブにおけるマイニング」(課題番号: 18700111) によるものです。ここに記して謝意を表すものとします。

参考文献

- [1] Yamamoto, Y., Tezuka, T., Jatowt, A., Tanaka, K., Honto? search: Estimating trustworthiness of web information by search results aggregation and temporal analysis, APWeb/WAIM, 2007
- [2] 熊本 忠彦, 灘本 明代, 田中 克己, 記事の印象を伝達するニュース番組生成システム wEE の設計と評価, 電子情報通信学会論文誌 D, Vol. J90-D, No.2, pp. 185-195, 2007