

情報分析システム WISDOM のユーザ評価とその分析

川田 拓也* 赤峯 享*[§] 河原 大輔^{†*} 加藤 義清* 乾 健太郎^{‡*} 黒橋 禎夫^{†*} 木俣 豊*

* 独立行政法人 情報通信研究機構, [§] 奈良先端科学技術大学院大学, [†] 京都大学, [‡] 東北大学

1 はじめに

Web が日常的になるにつれ Web 情報を利用した意思決定は珍しくなくなってきている。進学や就職活動において Web の活用は不可欠といわれる。日常的な購買活動においても我々は口コミや、専門家によるレビューなど Web 上の様々な情報を分析して意思決定を行っている。Web はテレビや新聞など既存のメディアとは異なり、誰もが自由に発信できるという特徴を持つ。一方で情報の匿名性の高さ故に、Web を情報源として利用する場合にはその信頼性が常に問題となる。現状では、人間が玉石混淆の Web 情報から有用かつ信頼できる情報を選び分けるには、高い判断能力が要求される。

このような背景から我々は、Web 情報を多角的な観点から組織化して提示しすることによってユーザの情報分析を支援する情報分析システム WISDOM (<http://wisdom-nict.jp/>) の開発を行っている。WISDOM は、入力される任意のトピックに対して関連するウェブページ 1,000 件を収集し、ページ発信者の解析、トピックの関連語解析、評価表現解析、主要・対立文解析等を行い、その結果を様々に統合してユーザに提供する (図 1)。

WISDOM の性能を評価するためには、ユーザの Web 情報分析における WISDOM の有用性を検証する必要がある。WISDOM の評価には三段階あると我々は考えている。まず第一に、WISDOM の各機能の定量的な性能評価をする段階で、次は我々の意図した情報の組織化が WISDOM によって実現できているか検証する段階である。その上で WISDOM の情報分析における有用性を検証する段階がある。しかし、実際には情報分析の有用性を測る指標を設けることは容易ではない。我々は第一段階である各機能の評価 [7, 5] に加えて第二段階、すなわち WISDOM がユーザに対して適切にわかりやすく組織化された情報を提供できたかどうかについて、既存の検索エンジンとの比較を通じたユーザ評価を行った。さらにユーザによる印象評定を行うことによって、間接的ではあるが、WISDOM の情報分析における有用性を検証した。

2 意思決定のための情報分析支援

情報分析システムの評価に当たり、まず、人間の情報探索、意思決定の過程について整理し、WISDOM がその過程をどのように支援すべく設計されているか述べる。

Web を用いて情報探索をする背景には、知りたい情報が明確で、それに見合う定まった答えを (多くは一度の検索で) 探索する場合と、目的が曖昧な状態で検索を繰り返しながら知りたい情報を明確にしていく場合がある [2, 6]。現実には未知の情報を検索し、知識を獲得してそこで終わるとは限らない。得られた情報を分析し、意

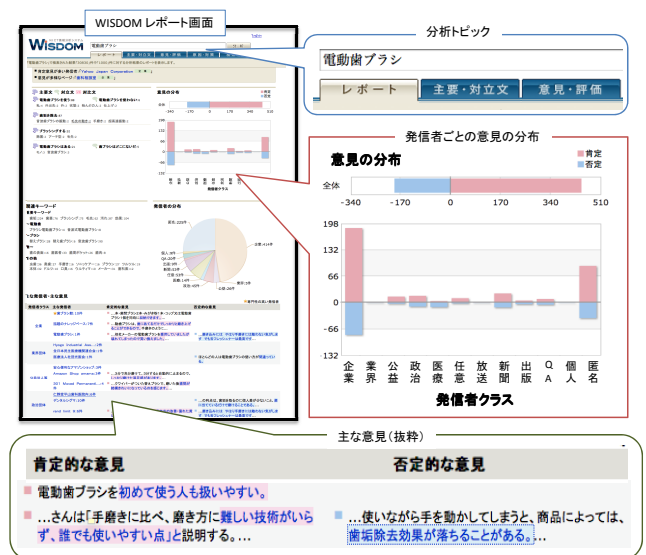


図1 WISDOM「レポート」画面

思決定に至ることもある。Simon ら [8] によると、意思決定はまず、新たな行為を必要とする状況の設定を経て、可能な行為の代替案を見出し、その中から選択する過程としてみなされる。Web による意思決定も同様に、その過程は図 2 のように捉えられる。図 2 を基にすると、各段階においては次のような支援があり得る。まず、段階 (0) は漠然とした興味はあるが意思決定をするには具体化されていない状況である。例えば漠然と「歯周病は怖い感じがよく知らない」と感じているユーザに対しては、「歯周病は歯垢が原因の一つである」といった百科辞典的な知識や周辺情報を提供することが有用で、Wikipedia や現状の検索エンジンがある程度担っている部分である。(1) に関しては、ユーザの願望や要求を満たす解決候補を提示することが有用であると考えられる。例えば、「歯垢を効率的に落としたい」というユーザに対しては「電動歯ブラシ」や「フッ素洗口」など様々な候補を提示する。この点については Web ページのクラスタリング技術 [4, 1] が有望である。(2) は利点や欠点のような、その解決候補の価値を示す情報収集が不可欠である。例えば、候補を「電動歯ブラシ」に絞って分析する場合は、「電動歯ブラシ」に関する Web 上からの多様な意見の抽出が支援内容として挙げられる。(3) では (2) で収集した情報の発信者や、全体に対する位置づけ (多数派か少数派かなど) を提示することによって、その情報の真偽判断を支援することができる。(2) で得た多種多様な情報をそれぞれ検証することは容易ではない。そこで (3) の観点から情報の集

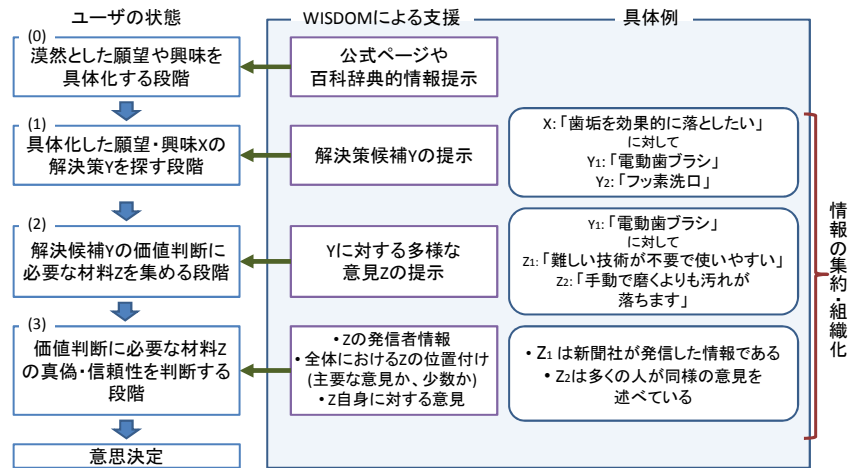


図2 意思決定の過程とその支援

約と組織化をして提示することも重要な支援過程となる。

図1は、WISDOMが分析した結果を集約して提示する「レポート」画面である。まず、(0)についてはWikipediaの定義文を表示することでそのトピックにおける一般的な知識を提供している。また、WISDOMにはトピックと関連するキーワードを表示する機能や、対策となる言明を提示する機能があるが、それによって(1)の状態にあるユーザの願望・要求を解決する選択肢提示としての機能が果たされる。WISDOMが特に重点的に支援しようとしているのが(2)、(3)である。例えば、「レポート」は意見が豊富で有用なページや、発信者ごとの意見の傾向の違いを表したグラフ、さらに代表的な意見などが提示される。図1でいえば、下部に示されている「商品によっては歯垢除去効果が落ちることがある」といった意見がリスト化される。それらを組織化することによって、ユーザの情報把握が容易になり、意思決定過程における(2)、(3)を支援する。

3 実験設計

本評価の目的は、我々の意図した情報の抽出と組織化がWISDOMによって実現できているか検証することにある。以下では実際にWISDOMを用いたユーザによる評価とその結果について述べる。

3.1 課題概要

実験協力者は100名で男性52名、女性48名だった。年齢構成は10-70代に渡り、20代が33%、30代が32%、40代が21%、50代が10%を占めた。職業は多様だった。募集の段階で、インターネットを週一度以上利用する人に限定した。なお、情報通信を専門とする社会人もしくは学生は1割強であった。

評価実験は、実験協力者がWISDOMを実際に操作しながら、我々が用意した質問項目に回答する方式を採った。実験協力者100名を募り、我々が構築したWebによる回答システムを通じて回答する方式とした。指定期間内であればどこでいつ回答しても良いこととした。ただし、実験協力者のWISDOMの操作履歴(分析トピックやクリックした箇所など)を記録しておき、実験終了後に分析を行った。

質問項目は記述項目、選択項目合わせて84項目を用意

した。質問内容については後述する。回答時間に制約は設けなかったが、WISDOMを実際に操作しながらの回答になるため、4-6時間を想定した。回答システムを回答の途中記録が可能な仕様とし、実験協力者は回答期間中いつでも回答を中断、再開できるようにした。一般に公開されているWISDOMは検索インデックスを日々更新しているため、日によって出力結果が異なる場合がある。そのため、ある時点のデータに固定した評価実験用WISDOMを別に構築した。データ以外は一般公開されているWISDOMと同様である。

回答項目は大きく分けて「実験協力者の属性」に関する質問と「既存の検索エンジンとの比較」、「WISDOMの総合的な評価」の3部から構成される。「実験協力者の属性」では、職業や年齢層の他、参考のためにインターネットの利用頻度や目的などを項目として設けた。残り2つについては次節で述べる。

3.2 比較課題

「既存の検索エンジンとの比較」は本評価実験の中心的な課題で、我々が用意した10トピックの中から実験協力者が選んだ2つと、実験協力者が別途自由に考えたトピック2つを実際にWISDOMと比較対象であるGoogleで分析しながら、質問項目に回答するというものである。以降では「比較課題」と呼び、我々が用意したトピックを「選択トピック」、実験協力者が別途自由に考えたトピックを「自由トピック」呼ぶこととする。選択トピックは以下10トピックである。

ホメオパシーの効果	成果主義は業績を上げる	任基ネット
ハイブリッドカー	コーヒーは体に悪い	抗がん剤
ジェネリック医薬品	マイナスイオン	ダムは必要か
赤ちゃんポスト		

比較課題で実験協力者が分析した計4トピックの質問内容はすべて同じである。100名のうち53名は常にWISDOMから先に分析し、残り47名は常にGoogleから分析した。WISDOMは目的や、操作方法に慣れるため、比較課題を行う前にWISDOMの操作マニュアルに従って一通り分析するよう指示した。

質問項目は次の通りである。「多様な意見の見つけれられたか」(以下「多様な意見」)、「情報発信者の立場の違いによる意見の相違が見られたか」(以下「立場による意見

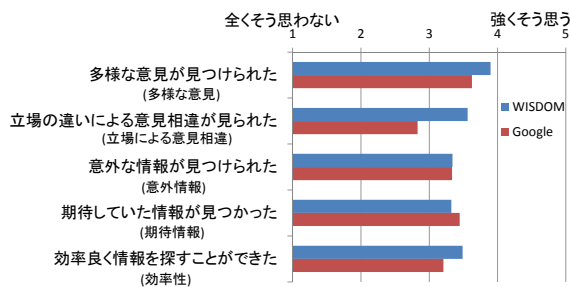


図3 WISDOM と Google の比較

相違)、「未知の意外な情報が発見できたかどうか」(以下「意外情報」)、「期待していた情報が確認できたかどうか」(以下「期待情報」)、「効率よく情報分析できたかどうか」(以下「効率性」)。各質問とも五段階評価とした。項目は選択トピック、自由トピック両方に共通する。

3.3 総合評価

「WISDOM の総合評価」は、WISDOM を一通り利用して実験協力者が感じた WISDOM の利点、欠点や利用法などについて問うものである。まず WISDOM の役に立った機能や不便だった点を挙げる項目を設けた。次に WISDOM がどのような場面／観点で役に立つか回答する項目を設けた。質問項目は次の 6 点で、同意～不同意までの 5 段階評価とした。(1)「何か迷っている時(ものを買う時など)に役に立つ(意思決定)」,(2)「Web 上から信頼できる情報を見つけ出そうとするとときに役に立つ(信頼情報発見)」,(3)「Web 上の口コミを見つけるのに役に立つ(口コミ検索)」,(4)「自分と同じ考えを持つ人を探すのに役に立つ(同じ立場の人検索)」,(5)「物事を様々な視点から見るのに役に立つ(多様な視点)」(6)「インターネット上の多様な情報を選び分け整理して把握するのに役に立つ(情報組織化)」。

4 結果

4.1 比較課題

WISDOM と既存の検索エンジン (Google) との比較項目の結果を図 3 に示す。図 3 は選択・自由トピックの各質問に対する全回答の平均値を表す。マン・ホイットニーの U 検定を行った結果、「多様な意見」と「立場による意見相違」および「効率性」に関する項目で WISDOM は Google を有意に上回り(いずれも $p < .01$)、「意外情報」と「期待情報」に関する項目では、WISDOM と Google には差が見られなかった。次に選択トピックか、自由トピックかによって WISDOM と Google の差に影響があるか調べた。選択トピックに限定した場合における WISDOM と Google の差異は、無条件で比較した場合と変わりなかったが、自由トピックに限定した場合は、「立場による意見相違」のみ WISDOM が Google より有意に高かった ($p < .05$) が、他の項目については有意な差異が見られなかった。今度は、WISDOM から先に使用した場合と、Google から先に使用した場合の違いについて調べた。常に WISDOM から先に分析した実験協力者の場合、WISDOM の方が Google より優れていた項目は「立場による意見相違」のみで ($p < .01$)、「期待情報」においては、Google の方が勝っていた ($p < .05$)。常に

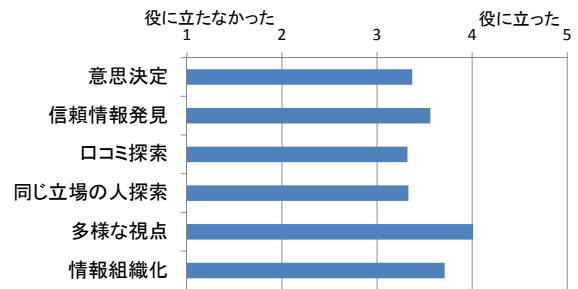


図4 WISDOM 総合評価

Google から先に分析した実験協力者の場合、全体の平均と同様に「多様な意見」と「立場による意見相違」および「効率性」(いずれも $p < .01$)に関する項目で WISDOM は Google を有意に上回っていた。すなわち、分析方法や分析する順番に関わらず、WISDOM は情報発信者の立場の違いによる意見の相違を見つけるのに有用であることが分かる。また、多様な意見を発見するのにもある程度有用だった。意外な情報や期待した情報などユーザの興味を引く特定の情報を提供する場合においても通常の検索エンジンと差異はないことが確認された。

4.2 総合評価

総合評価の中で「WISDOM の役に立つ点」の回答結果を図 4 に示す。グラフの示す点数(横軸)は実験協力者全員の回答の平均値を表す。グラフが示すように、「多様な視点から見ると」という点、「情報の組織化ができる」、「信頼できる情報を発見できる」という点が特に評価されている。一方で、「口コミを見つける」といった点や、「同じ立場の人を見つける」といった点では相対的に評価されていなかった。WISDOM の有用性とは逆に、WISDOM を使用していて不便だった点を自由記述で回答する項目も設けた。自由記述回答を分析した結果、実験協力者が回答した WISDOM の不便な点は以下三つに集約できることがわかった。(1) ユーザビリティ。(2) 検索結果の数とクエリとの関連性(検索精度)。(3) 意見・評価の解析誤り(意見精度)。そこで人手で回答を分析し、上記 3 点に加え、「不便な点がない」という回答と「その他」の 5 項目に分類した。最も多かったのが「ユーザビリティ」に関するもので、42% を占めた。例えば「情報過多でどこに注目したらよいかわからない」、「操作性が悪い」などの意見が見られた。次に「検索精度」に関する意見が多く、23% を占めた。「検索結果が少ない」、「クエリとは無関係の検索結果が目立った」といった意見が見られた。「意見精度」に関しては 20% を占め、「極性の誤りが目立つ」という意見がほとんどであった。その他の意見としては「意見がページ(文章)単位ではなく、文単位で切り取っていることは問題ではないか」「使う機会が限られる」といった意見が見られた。

4.3 自由トピックの分析

自由トピックと WISDOM の評価の関係を見ることによって、ユーザの動機や分析課題の性質が評価に与える影響について述べる。

まず、WISDOM の評価が高かったトピックを表 1 にまとめた。比較課題の各質問項目の合計点数の高い順に 5 トピック選択した。左の「絶対評価」は、比較課題に

表1 WISDOM の評価が高かった5トピック

WISDOM評価	絶対評価	相対評価
高	ミックス犬	普天間基地辺野古移転
	学校選択制度	ミックス犬
	自転車専用道路	学校選択制度
	ブラウン管テレビの画質	尖閣諸島 日本領土
低	メタンハイドレート	ブラウン管テレビの画質

表2 WISDOM の評価が低かった5トピック

WISDOM評価	相対評価	相対評価 (ヒット件数:1,000件, 意見:500件以上)
低	coron	ソーラーパネル
	トヨタリゲントガムシ	スマートフォン
	coron ロボット	エコカー減税
	島田紳助 どう思う	東京ディズニーリゾート
高	おいしいお店ランキング	アニマルセラピー

おける WISDOM に対する質問項目の合計点数が高いものから5トピック抽出したものである。「相対評価」は WISDOM の評点が高く、Google の評点が低かった5トピックを抽出したもので、WISDOM の評点と Google の評点の差分が大きいものから順に並べている。その結果、賛否両論ある政治・社会問題や利用物に関するトピックは比較的评价が高い結果となった。一方で、表2は Google と比較し相対的に WISDOM の評価が低かったトピックである。検索ヒット件数や意見の数が十分ではなく分析に適さなかったことが要因の一つである。表2の右側のカラムは検索ヒット件数が1,000件以上、意見が500件以上返ってきたトピックに限定したものである。自由トピックの分析動機を調べると、「エコカー減税」と「スマートフォン」については「CMでよく宣伝しているから」「最近のトレンドだから」という動機であった。図2でいえば、「(0) 漠然とした興味」の状態での分析を始めたと思われる。むしろ WISDOM は(2)、(3)のサイクルを支援することを意図しているため、賛否両論ある政治・社会問題や利用物に関するトピックが高い評価を得られたのは我々の意図に沿った結果であると考えられる。

5 考察

比較評価において意外な情報や、ユーザが期待していた情報の発見については既存の検索エンジンと差がなかった。ユーザの興味を引く特定の情報を提供する部分については既存の検索エンジンに引けを取らず、さらに多様な意見の発見や、発信者の立場の違いによる意見の相違を見つけることに有用であったことは、我々の意図するユーザの真偽判断に必要な情報の組織化に成功しているといえ、図2における(2)、(3)の支援が適切に機能しているといえる。総合評価において「多様な視点からの俯瞰」「情報組織化」、「信頼情報発見」という点が評価されていたため、印象評定ではあるが、総合的に見て WISDOM を利用した情報分析に有用であったことも示唆される。

今後は(0)、(1)の支援も含めた統合的評価が課題になる。動機が曖昧なユーザの願望、要求を具体化し、さらにその解決候補の絞り込みを支援する機能を統合することで、ユーザの意思決定にどのように影響を与えるか評価することが意思決定の過程全体を視野に入れた情報分析システムの評価につながるだろう。

6 関連研究

Web 情報の分析をする技術には様々な分野が関わりを持つ。まずキーワードを基に Web 情報の組織化して提

示する技術としては、Clusty (<http://clusty.com/>) など商用サイトがすでに公開されている。この点に関して WISDOM は Web ページをより深く分析し、意見や発信者といった観点から組織化する。同様に深い Web ページの分析技術としては、Dispute Finder [3] が挙げられる。Dispute Finder は Web から論争になっている主張を抽出およびデータベース化し、それを基に Web ページに書かれた主張について、その反論となる情報を表示するシステムである。WISDOM は各意見とその反対意見だけではなく、図1の「レポート」のように、Web 情報の全体把握が可能な形で提示する。

また、Web で得た情報の真偽判断支援については Web ページの信頼性判定技術がこれまでもいくつか提案されている。例えば、Wassmer ら [9] や Weerkamp ら [10] などが記事の長さなどページの外観的側面を信頼性の指標とみなして Web ページやブログの信頼性の自動推定を行っている。それに対して、WISDOM は図2における(2)、(3)を重点的に支援し、ユーザが真偽判断の際に有用な情報を整理して提示する立場を取る。

7 結論

本稿では WISDOM の評価実験を通して、WISDOM が人間の情報分析行動に沿った形で必要な情報を提示し、支援できることを示した。さらに、分析動機が曖昧な状態での支援など、WISDOM の情報分析支援という観点からの課題を整理した。

参考文献

- [1] 馬場康夫, 新里圭司, 柴田知秀, 黒橋禎夫: キーワード蒸留型クラスタリングによる大規模ウェブ情報の俯瞰, 情報処理学会論文誌, Vol. 50, No. 4, pp. 1399-1409 (2009).
- [2] Broder, A.: A taxonomy of web search, *ACM SIGIR Forum*, Vol. 36, No. 2, pp. 3-10 (2002).
- [3] Ennals, R., Trushkowsky, B. and Agosta, J.: Highlighting disputed claims on the web, *WWW2010*, pp. 341-350 (2010).
- [4] Ferragina, P. and Gulli, A.: A personalized search engine based on web-snippet hierarchical clustering, *WWW2005*, pp. 801-810 (2010).
- [5] Kato, Y., Inui, K., Kurohashi, S. and Shibata, S.: Identifying the information sender configuration of web pages, *Proc. of the 2009 IEEE/ACM/WIC International Conference on Web Intelligence (WI'09)*, pp. 335-340 (2009).
- [6] Marchionini, G.: Exploratory search: from finding to understanding, *Communications of the ACM*, Vol. 49, No. 4, pp. 41-46 (2006).
- [7] Nakagawa, T., Inui, K. and Kurohashi, S.: Dependency tree-based sentiment classification using CRFs with hidden variables, *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 786-794 (2010).
- [8] Simon, H.: *The New Science of Management Decision*, Prentice-Hall, Englewood (1977). [稲葉元吉・倉井武男(訳) 1979 『意思決定の科学』 産業能率大学出版部].
- [9] Wassmer, M. and Eastman, C.: Automatic evaluation of credibility on the Web, *Proceedings of the American Society for Information Science and Technology*, Vol. 42, No. 1 (2005).
- [10] Weerkamp, W. and de Rijke, M.: Credibility Improves Topical Blog Post Retrieval, *Proceedings of ACL-08: HLT*, Columbus, Ohio, pp. 923-931 (2008).