

Web ブラウジング履歴からの自己の興味の理解

村上 晴美

(大阪市立大学 学術情報総合センター)

Keywords: 自己, 興味, WWW

1. はじめに

本研究は、人間が何に興味を持っているかを知るためのツールの開発を目的とする。本研究では、(1)「人間は自己の興味に従って Web ブラウジングを行う」、(2)「個人の Web ブラウジング履歴から計量的に個人の興味を表す情報を抽出できる」、(3)「個人の Web ブラウジング履歴から計量的に抽出した語の提示が個人の興味の理解に役立つ」という仮説をたて、Web ブラウジング時に抽出した語と URL を興味空間と呼ぶ 2 次元空間上に時系列に配置することにより個人の興味の理解を支援するシステムを試作した。本稿ではシステムと、新聞記事の Web サイトを対象としたシステムの予備的実験の概要を述べる。詳細は文献を参照されたい。

2. システムの概要

本システムは(a)ユーザが Web ブラウジングを行うオーバーレイ Web ブラウザと、(b)ユーザの興味空間を表示する興味空間ブラウザから構成される。オーバーレイ Web ブラウザでは、ユーザが Web ブラウジングを行う毎にクリックしたテキストであるリンクテキストから「(1)ひらがなと特殊文字以外が 2 つ以上続いた文字列を抽出し、(2)ある条件にあてはまる不要語を削除する」という手法を用いて語を自動的に抽出する。抽出された語は Web ページを閲覧した日付及びリンク先の URL と組で索引として保存される。興味空間ブラウザでは、索引の中から 2 回以上出現した語の共起に基づき数量化 3 類を用いて、語と、URL を表す Web アイコンを 2 次元空間上に表示する。日付情報を利用して興味空間の変化を時系列に見ることができる。図 1 に興味空間ブラウザの画面例を示す。

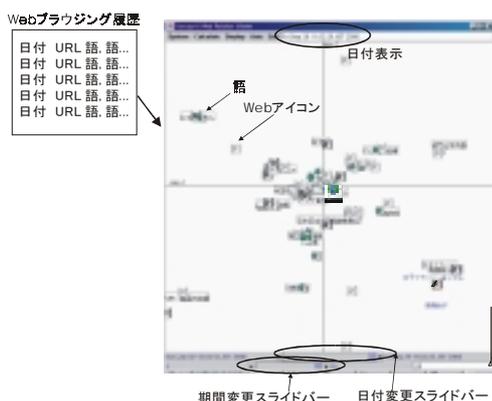


図 1 : 興味空間ブラウザの画面例

3. 実験

新聞記事を対象とした Web サイトにおける語の抽出手法とシステムの有用性について検討する。

3.1 方法

被験者は 32、37、37、24 歳の女性 4 名であり、順番に A、B、C、D と呼ぶ。Web 利用経験はそれぞれ 2、3 年程度である。毎日インタラクティブの Web サイトから HTML ファイル (記事) をダウンロードして、オーバーレイ Web ブラウ

ウザで記事の見出し一覧を表示して、見たい記事を自分のペースで 1 日につき 1 つ以上見るよう指示した。ブラウジング後に履歴を回収し、抽出された語が語として妥当かどうか筆者が判定した。次に語が「興味語としてふさわしいか」どうか被験者に評定させた。最後に興味空間ブラウザを使って、語のみ表示するテスト 1 と、語と Web アイコンを表示するテスト 2 を行い、それぞれ「興味空間を表しているか」「自分を知るのに役立つか」「過去の想起に役立つか」などの質問に対して被験者に評定させた。評定は 5 段階 (5 : 非常に思う、4 : やや思う、3 : どちらともいえない、2 : あまり思わない、1 : 全く思わない) で行った。

3.2 結果と考察

(1) 見出しから抽出された語の評価

新聞記事の場合リンクテキストは見出しに相当する。一つの見出しに対して平均 3.62 個の語が抽出された。被験者 A の履歴から抽出された語 416 個が語として妥当かどうか筆者が判定したところ、妥当 384 (92%)、修正すれば妥当 28 (7%)、妥当でない 4 (1%) と、高い精度で語が抽出できた。各語につき興味語としてふさわしいか被験者が自己評定した結果、高頻度語 10 語に関して平均 3.79 であった。この質問で被験者が 5 または 4 と評定した語の割合を興味度適合率として算出したところ高頻度語 10 語について 63% であった。新聞記事を対象とする場合には高頻度語が被験者 (自己) の興味を表すことが示唆された。

(2) システムの評価

アンケートの結果、テスト 1、2 とともに、システムは被験者の興味空間を表しており (テスト 1: 平均 4.25、テスト 2: 平均 4.00。以下同様)、過去の想起に役立つ (4.25、4.25) ことが示唆された。自分を知ることに役立つ (4.00、3.25) かは、やや肯定的な結果である。総じて、テスト 1 よりもテスト 2 の方が評価が低いと、表示される語と Web アイコンの類似度が必ずしも対応していないことが理由である。全体としてシステムの有用性が示唆されたが、表示や操作方法に関する問題点が指摘された。

4. おわりに

Web ブラウジング履歴を利用して個人の興味の理解を支援するシステムを試作した。新聞記事の Web サイトを対象とする予備的実験を行ったところ、(1)語の抽出手法が妥当であること、(2)高頻度語が被験者の興味を表していること、(3)システムが自己の興味空間を表しており、過去の想起や自分を知ることに役立つ可能性があることが示唆された。今後は語の抽出・提示手法やユーザインターフェースを改善するとともに、Web での評価実験を進める必要がある。

文献

(村上, 平田 2001) WWW からの情報の獲得・整理支援 - 思考・興味空間ブラウザ -, 情処研報, 2001, 20, 167-174, 2001.

(MURAKAMI Harumi)