

ユーザの興味に基づく記事注目度を用いた RSS 記事閲覧支援システム

浅野 光

Web サイトは日々更新され、多量の新しい情報をユーザに提供している。多くのニュースサイト、ブログなどではサイトの更新情報の配信に RDF Site Summary (以下、RSS) を利用している。RSS はサイトの見出しや内容の要約を記事として XML で記述するフォーマットである。RSS でサイトの更新情報を記述し、配信するサービスを RSS フィード (以下、フィード) と呼ぶ。フィードを RSS リーダというソフトウェアに登録することにより複数のサイトの記事をまとめて閲覧できる。しかし登録フィード数が増えると取得する記事が多くなり、それらの記事にはユーザの興味のない内容のものも多く含まれる場合があるため、その中から興味のある記事を探し出すには多大な手間がかかる。

そこで本研究では、多量の記事の中からユーザが興味を持ちそうな記事を推定し、それらを優先して表示することで、記事閲覧の効率化を図るシステムの開発を行った。ユーザの興味を得る手掛かりとしてユーザの操作から、記事の評価と、どの単語が含まれた記事を優先的に見ているかという情報を得る。記事の評価として各記事についての未既読状態、有用、不用という評価、読まずに既読状態にされたか、という情報(以下、フラグ情報)を得る。単語については記事に含まれる単語に重みを付け、キーワードとして抽出し、単語ごとに閲覧回数、出現回数、最終閲覧日時、最新出現日時を得る(以下、単語閲覧情報)。フラグ情報から記事を興味あり、興味なしの二種類に大別する。興味ありの記事のキーワードの重みを正の値、興味なしの場合は負の値とし、単語ごとにキーワードの重みを積算した値を求める。その積算値と単語閲覧情報を利用し、単語に対する興味の度合いを表す単語重要度を求める。そして各記事中のキーワードの単語重要度の合計値を、記事に対する興味の度合いを表す記事注目度とする。

記事注目度の有効性を確認するために、被験者にシステムを利用してもらい記事注目度を付与する実験を行った。新着記事を記事注目度順に並び替え、その上位にユーザが有用、既読とする記事がどれだけ含まれたかを評価した。実験では大学生 3 名に 250 件の記事を閲覧してもらいデータを収集した。その結果、新着記事を記事注目度順に並び替えた場合では有用、既読の記事が 1 割~4 割しか推定できず、ユーザの興味の推定が十分行えなかった。しかし既にフラグ情報を付けた記事を記事注目度順に並び替えた場合では 9 割の有用、既読記事が上位に入っていたため、ユーザの興味の学習はできていることが分かった。学習ができているのに推定が不十分だった原因として、キーワードの抽出が適切に行われていないことが考えられる。その結果、キーワードの単語重要度によって決まる記事注目度が、ユーザの興味とずれてしまったものと考えられる。このことから、より正確に記事の主題を表すようなキーワード抽出手法の開発が残された課題として考えられる。

(指導教員 阪口 哲男)