

Web サイトのお知らせページにおけるユーザーの閲覧履歴と記事の更新日による興味度の視覚化

Visualization of interest level based on user browsing history and article update date on the website's announcement page

東京工科大学 戦略的教育プログラム (Cloud Native Dojo)

有賀皓哉¹⁾, 川端ももの¹⁾, 田中美帆¹⁾

指導教員 串田高幸²⁾, 研究協力者 高橋風太²⁾

1) 東京工科大学コンピュータサイエンス学部 3 年生

2) 東京工科大学大学院バイオ・情報メディア研究科

キーワード: Web サイト, コサイン類似度, 興味度, TF-IDF

1. はじめに

Vaundy の公式サイトのお知らせページは CD の発売情報やライブの開催予告を告知する記事が掲載されている¹⁾。これらの記事は、更新日が新しい順に並んでおり、順次追加されていく。ユーザーが記事閲覧するまでに複数の動作がある。例えば、Vaundy のライブ情報を知りたいとき、Vaundy の公式サイトのお知らせページを開く。次にタイトルをから、ライブ情報が載っている記事を探しクリックする。その際、興味のある記事を見つけるためにはタイトルを読む必要がある。

コサイン類似度はユーザーの興味を数値で表すために使用される[1]。コサイン類似度は比較対象の 2 つがどのくらい似ているかを比較する。

2. 課題

ユーザーが再度記事閲覧したいとき、追加された記事に埋もれてしまうため短時間で探すことができない。また、ユーザーは記事の内容を判断するために長いタイトルを読む必要がある。ユーザーの興味がある記事が見つかるまで読む必要があるため、時間がかかる。さらに、ユーザーは興味がある記事であるかをタイトルで判断する必要がある。記事のページに遷移し数行読んだ後に、興味の無い記事であると判断する場合がある。このとき、お知らせページと記事のページ間の遷移に時間を要する。

3. 提案

ユーザーの興味がある記事を選択するために興味

度という指標を作成する。ユーザーの興味を引く記事にするため、記事を鮮やかな魚として表示する。魚がページ内で見やすい数が 12 匹であるため、ページには興味度の高い記事 12 件を表示している。ユーザーが認識しやすい数を算出するスタージェスの公式から、表示する魚が 12 匹のとき魚は 5 種類となる[2]。ユーザーの興味度によって表示する魚を変更する。表示する魚は大きさ順に興味度が高くなり、シャチ・ナポレオンフィッシュ・タイ・クマノミ・ミジンコの 5 種類である。記事を魚にすることで、ユーザーは自分の興味がある記事をタイトルを読まなくても魚の種類で判断することができる。提案するページをお知らせページとする。

興味度の計算方法を以下に示す。(1)記事閲覧した人は閲覧した記事に興味がある[3]。そのためユーザーが過去に閲覧した記事の本文の名詞をリスト化する。(2)品詞の中で名詞が最も閲覧者の興味が強いため、記事が更新されたときに記事の本文は形態素解析により名詞のみリスト化する[4]。全ての記事の名詞リストを抽出する。(3)TF-IDF を使用し出現頻度を計算する[5]。(1)のリストと(2)の各リストの類似度をコサイン類似度を用いて算出する[1]。(4)経過日数の重みを計算する。最も古い記事を 0.0 とし、最新の記事を 1.0 とする。更新からの経過日数の重み付けを図 1 に表す。(5)記事の内容がライブ開催を例とする特定の日時での内容であるとき、時間経過で記事への関心が下がっていく[3]。古い記事の興味度を下げするために、(3)のコサイン類似度の計算結果に記事の更新からの経過日数

¹⁾ Vaundy . "Vaundy Official Website" .

<<https://vaundy.jp/news/1/>>. (参照 2023-10-19)

の逆数を計算する。(6)(5)の計算結果を(2)の記事に掛け興味度を計算する。興味度の上位 12 件がお知らせページに表示する記事である。

表示する記事の興味度の最大値と最小値から公差を求める。公差の 20%ごとに表示する魚を変える。魚の一覧を図 2 に表す。

ユーザーが記事に遷移したとき、ユーザーの名詞リストに遷移した記事の名詞が追加される。

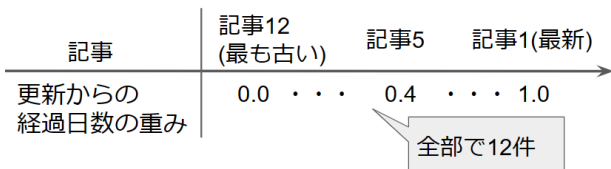


図 1. 更新からの経過日数の重み付け

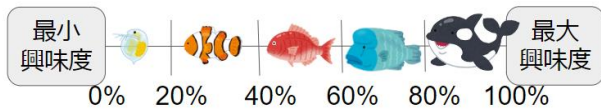


図 2. 興味度で表示する魚を変える

4. デモンストレーション・シナリオ

お知らせページの魚はユーザーごとの名詞リストに格納されている名詞によって興味度が計算され決定されている。提案するお知らせページを図 3 に表す。お知らせページの魚をクリックすると、魚と対応するタイトルの記事のページに遷移する。興味度は、ユーザーが記事をクリックして遷移した際に計算され更新する。お知らせページに戻ると新たに計算された興味度に基づいて魚の表示が変更される。その際、計算前と魚の表示が異なる記事は updated と魚の下に表示される。web 上で記事を読覧しお知らせページに戻ると、名詞リストに遷移した記事の本文の名詞が追加されている事を見せる。名詞リストを図 4 に表す。

興味度はユーザー毎に計算するためユーザーによって表示される記事や魚が異なる。ライブの記事を多く見る人はライブ関連の記事がシャチやアンコウで表示されるが、出演番組の記事を多く見る人は番組情報の記事がシャチやアンコウで表示される。

記事を魚で表示し魚の種類で興味の高さを表す数値を絞ることで、記事の題名を読まずとも興味のある記事を判断できる。そのため記事のタイトルを読む時間を削減し、記事が選びやすくなる。

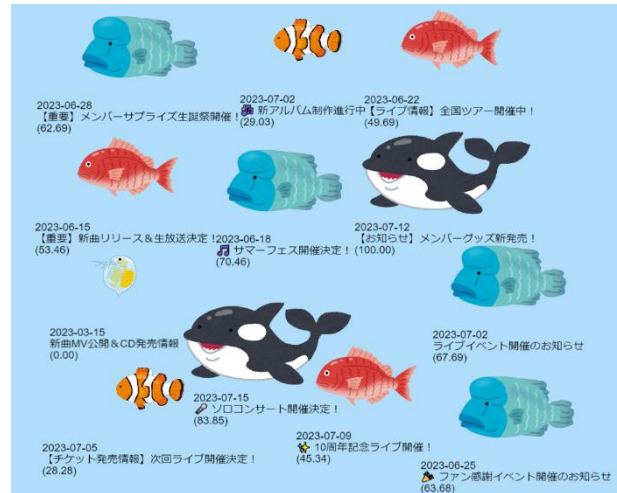


図 3. 提案するお知らせページ

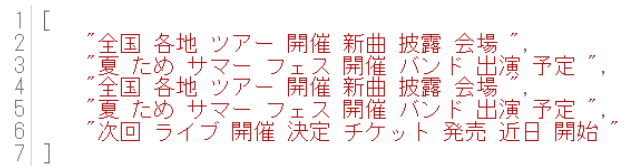


図 4. お知らせページの名詞リスト

謝辞 提案するソフトウェアのフロントエンド開発にご協力いただいた東京工科大学コンピュータサイエンス学部先進情報専攻の近藤悠斗さんに御礼申し上げます。

5. 参考文献

[1] 新原 俊樹. コサイン類似度を用いた行政文書タイトルの同一判定. About the journal. 2022. vol82. p.61-63

[2] David W. Scott . Sturges’ rule . WIRES Computational Statistics Volume 1, Issue 3. Nov 2009. p.261-367

[3] 寺菌 浩平, 楠嶺 生宏, 鳥谷 彰, 松岡 直樹, 駒場 祐介, 柳沼 義典: ユーザーの関心度の経時変化を考慮したコンテンツレコメンド技術, マルチメディア, 分散協調とモバイルシンポジウム 2158 論文集, Vol. 2020, pp. 963–969 (2020).

[4] 沢井康孝, 山本和英. 文書に対する大衆の興味の強さの推定. 自然言語処理. 2008 年 15 巻 2 号.p.101-136

[5] Michel Capelle, Flavius Frasinca, Marnix Moerland, and Frederik Hogenboom. 2012. Semantics-based news recommendation. In Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics (WIMS '12). Association for Computing Machinery, New York, NY, USA, Article 27, 1–9.