6S-2

タグクラウドを用いた注目情報提示方式

吉田　慶章 †　　柿崎　淑郎 ‡　　辻　秀一 ††

† 東海大学電子情報学部　　‡ 東海大学連合大学院理工学研究科　　†† 東海大学情報理工学部

1　はじめに

IT 技術の進歩と共に，私たちは Web との共存をしている．情報発信者側の提供する情報が主体であった以前の体制 [3] から，ユーザ自らが情報を発信し貢献していく時代，言わば Web2.0 [1] の世界に突入している．その結果，Web 上には多種多様な情報が氾濫し，その中からユーザの要求に適合した情報を探し出すことが困難な状況にある．さらに，検索によって情報を得るためには，得ようとする情報に関連したキーワードを思考する必要があるため，ユーザが無知である内容に関する情報のキーワードは発想することができず，情報を探し出せないという問題点がある．

そこで「ユーザが自ら検索をすることなく情報を得る」という切り口から，ユーザの知識に依存しない方式で，Web 上から情報を視覚的に提供する方式として，タグクラウドを用いた注目情報提示方式を提案する．本提案方式によって，ユーザは自己の知識に依存せず，膨大な情報の中から注目されている情報，ユーザに有益な情報を視覚的に得ることが可能となる．

2　関連方式

ユーザが Web から情報を取得する手法として，検索エンジンでのキーワード検索や，Web サイトから提示された情報からユーザの主観で取捨選択を行うなどが挙げられる．また，閲覧先の Web ページのリンクや広告を辿る事で思い掛けぬ情報を得ることもある．このように，従来方式では，ユーザは情報を得るために最低限のアクションを自ら起こす必要があった．しかし，前述したようにユーザが無知である情報に関してはキーワードの発想ができないことから，能動的なアクションのみでは，知り得ない情報が存在することがわかる．

ニュースサイトなど，時間に基づいた表示方式では，重要度に関係なく新着情報が常に上位に表示されてしまう．ランキング形式での情報提供方式では，人気情報が常に上位に表示されるため，下位情報を得るには，ページの下位部分までスクロールする，または，別ページに飛ぶなどのアクションが必要であった．さらに，その中から主観に基づいた情報の取捨選択をするため，本当に注目されている情報を見逃してしまうことが考えられる．そこで，ユーザの能動的なアクションを要せず，注目されている情報を提供することが必要であると考えた．

情報の提供方式として，今回タグクラウドを用いる．タグクラウドとは，タグの一覧を頻度に応じて表示する方法 [2] を言う．ここで，代表的なソーシャルブックマークである “del.icio.us” *のタグクラウドを図 1 に示す．

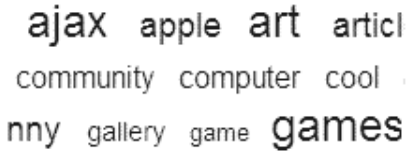

図 1: del.icio.us のタグクラウド

3　提案方式

従来方式の問題点を解決するため，ユーザに能動的なアクションをさせずに情報を視覚的情報として提示する手法として，タグクラウドを用いる本方式を提案する．

3.1　処理フロー

まず，指定した複数のニュースサイトの RSS フィードを取得する．複数の RSS フィードを取得することで，取得情報の分野の偏りをなくし，注目するべき情報を明確に得ることができる．また，ユーザの知識にも依存せず情報を集合知 [1] として取得することが可能となる．その情報に対し，正規表現による不要語除去，形態素解析による品詞分類を施し，名詞のみを抽出する．抽出した名詞ごとの出現回数を抽出日時別にカウントし，データベースに各要素を揃えて格納する．なお，キーワードの重み付けは，日時別の出現回数に減衰指数関数をかけることで実現している．各キーワードの重みからフォントや色調に変化を付け，重みが規定値を超えている情報をタグクラウドで表示する．そして，予め設定した更新間隔で RSS フィードを読み込むことで，データベースとの比較により情報の差分を取り出し，同様に正規表現からの処理を繰り返す．また，本研究での指標は時系列に沿った出現回数と減衰指数関数の積であり，より注目度の高いキーワード，新鮮度の高いキーワード，初出現のキーワードに重く

A Method of Information Presentation using TagCloud
†Yoshiaki Yoshida　‡Yoshio Kakizaki　††Hidekazu Tsuji
†School of Information Technology and Electronics, Tokai University
‡Graduate School of Science and Technology, Tokai University Unified Graduate School
††School of Information Science and Technology, Tokai University

*http://del.icio.us/tag/

掛かるようになっている．本方式では表示行の間隔は均一にする．この結果，フォントの大きさにより，キーワードが相互に重なり合うことが考えられるが，より重みの高いキーワードを強調するため，キーワード同士の重なりは，タグクラウドの重要な要素の 1 つであると考える．本提案方式の処理フローは以下の通りである．

1. ニュース情報を取得する
2. 正規表現を用いて，取得した情報から不要な文字列を除去する
3. 形態素解析を用いて，取得した情報を品詞分類し，名詞のみを抽出する
4. 抽出した名詞の出現数を用いて，キーワードの重みを算出する
5. キーワードの重みに従った情報の提供をタグクラウドによって行う
6. 過去に取得した情報との比較を行い差分を取り出し，1 からの処理を繰り返す

3.2 重み付け

ここで，上記の減衰指数関数を定義する．ある時間 t_n におけるキーワード k の出現回数を $v_{t_n,k}$，ある時間 t_n における減衰指数関数を $e^{-\alpha t_n}$，当日を 0 とした日数を n．ただし，$0 \leq n \leq n_{max}$．その結果，キーワード k の重み w_k は以下のように表される．

$$w_k = \sum_{i=0}^{n_{max}} v_{t_n,k} e^{-\alpha i} - \frac{\sum_{i=1}^{n_{max}} v_{t_n,k}}{n_{max} - 1}$$

4 実装と評価

4.1 実装

実装は，ニュースの取得に RSS フィード，形態素解析に MeCab，DB に MySQL を利用した．開発言語は PHP で行った．今回実装したタグクラウドを用いた情報提供画面例を図 2 に示す．

図 2: 情報提供画面例

4.2 評価

ここで，従来方式と本提案方式を相対的に評価する．新着情報が常に上位に更新されるニュース形式の表示方式に対し，新着情報の全てが重要であるのかという問題点が存在する．本提案方式では新着情報から過去の情報までをデータベースに格納してあるため，両者を比較した結果から，キーワードごとの数値化した重みをフォントの大きさ，色調に対応させている．表示キーワードのフォントの大きさにより，現時点での最注目キーワードを，色調により，出現キーワードの新鮮度を，双方の指標から視覚的情報として提示することが可能となり，この問題点を解決できていると考えられる．

次に，ランキング形式など人気情報が常に上位に表示される方式に対し，下位情報を得るには何らかのユーザのアクションを要するという問題点を検証する．表示位置に関して，設定した間隔でのデータ取得毎に，キーワードの重みに依存せず，キーワードをランダムな位置で表示させている．この結果，同一画面にて注目度の大小に捉われず，キーワードを知ることが実現されている．よって，この問題点を解決できていると考えられる．これらの評価から，注目情報を視覚的情報としてユーザに提供することで，従来方式の問題点が解決できていることがわかる．

5 まとめ

本稿ではタグクラウドを用いた注目情報提示方式を提案した．検索によって情報を得るためには，得ようとする情報に関連したキーワードを発想する必要があるため，ユーザが無知である内容に関して情報を探し出せないという問題点が存在する．また，提示された情報から取捨選択をする場合，新着情報が上位に更新される表示方式では，新着情報を得やすくなる問題点，ランキング形式による表示方式では，上位人気の情報を得ることが多くなるなど各項目に対応する問題点が存在した．ユーザが自ら検索することなく情報を得るという切り口から，氾濫した情報の中から，注目されている情報，ユーザに有益な情報を，上記の問題点を解決し，視覚的情報として的確に提供することを実現した．今後は，取得する情報源の評価，より良い重み付け関数の模索を行い，精度の高い表示を実現したい．

参考文献

[1] T. O’Reilly. What Is Web 2.0, 2005.

[2] 大向一輝．Web2.0 と集合知．情報処理, Vol. 47, No. 11, pp. 1214–1221, 2006.

[3] 橋本大也．Web2.0 とは何か．情報処理, Vol. 47, No. 11, pp. 1195–1204, 2006.