

# ニュース情報のための視覚化インターフェイス

野間田 佑也\* 星野 准一\*\*

多摩美術大学\* 筑波大学大学院システム情報工学研究科\*\*

本稿では、頻繁に更新されるニュースサイトやブログサイトで配信される RSS のための視覚化インターフェイスを提案する。新着情報のみを逐次読んでいく従来のフィードリーダーに対して、どういったキーワードが多くサイトや記事で取り上げられ話題になってきているのかといった、全体の様子や時間的変化の把握を支援することを目的とする。本システムでは、登録された複数の RSS から取得した記事の形態素解析によりキーワード抽出を行う。キーワードの発生からの経過時間や各記事における使用回数、他のキーワードとの共起関係を視覚的にユーザに提示することで、動的情報の閲覧と理解の支援を行う。

## Visualization Interface for Information Exploration and Observation of News

Yuya Nomata\*, Junichi Hoshino\*\*

Tama Art University\*

Graduate School of Systems and Information Engineering, University of Tsukuba\*\*

**Abstract** - This paper propose a novel visualization system for exploration and observation of news articles. The goal of the system is to support users in exploring and observation. The system provides visual summary of news articles.

**Keywords:** information visualization, information exploration, browsing, news, RSS

### 1. はじめに

現在、ニュースサイトやブログサイトをはじめとする多くのサイトが日々絶えず更新しており、情報収集を目的として継続的に複数のサイトの更新状況を確認することが日常的になってきている。しかしながら、確認対象となるサイトの数が多くなるにつれて、頻繁に更新するサイトや、一方でいつ更新する分からないサイトの更新の有無や内容を絶えず監視しつつ、現在どういった事柄が多くサイトや記事で取り上げられ話題になってきているのかといった、全体の様子や時間的変化を考慮して把握するのは困難である。

こうした、ユーザの更新作業の負担を軽減するため、更新情報を、フィードリーダーと呼ばれるアプリケーションで読み込める RSS(Really Simple Syndication)<sup>1</sup>と呼ばれる文書フォーマットによって配信するサイトが増えてきている。フィードリーダーは自動的に登録された RSS の内容を取得し、ユーザに提示することで、ユーザは自ら各サイトへ訪れることなく、更新の有無と内容を一元的に確認することができる。しかしながら、フィードリーダーで主流となっているインターフェイスは、メーラーのようにタイトルをリストとして提示するメーラー型や、電光掲示板のように、新着記事のタイトルのみを順

番に提示することで、ユーザが受動的に情報を得ることができるティッカー型があるが、全体の傾向を把握したい場合には、ユーザ自身が新規に配信されるすべての記事に目を通して、ユーザ自身が判断を行う必要がある。

そこで本稿では、情報視覚化によって、RSS の更新状況および、配信される記事におけるキーワードの出現状況をユーザに視覚的に提示することにより、ユーザが、動的情報の更新の確認を容易に行うことができ、かつ、どんな事柄が話題になっているのかといった全体の様子を知ることができるようにするための継続的な利用を想定したインターフェイスの開発を目的とする。

### 2. 関連研究

ニュース情報を視覚的に提示するインターフェイスの研究が行われている[1]。これは、各記事の発生日時からの経過時間を波紋のメタファを利用することにより視覚的に提示し、広がる角度によってカテゴリを提示し、ユーザが興味のあるカテゴリかつ最新の記事かつアクセスを容易にしている。この方式が、波紋表現により記事のカテゴリと発生からの経過時間を提示するのに対し、本手法では、個々の記事が対象ではなく、複数の情報源全体の概要を視覚的に提供することに焦点を当てている。

データベースからの情報探索のための情報視覚化に関しては、これまで多くのシステム開発や研究が行われてきた[2][3][4]。これらの多くは、静的なデータベースを対

<sup>1</sup> 他に、Rich Site Summary, RDF Site Summary の略語とされる場合がある。

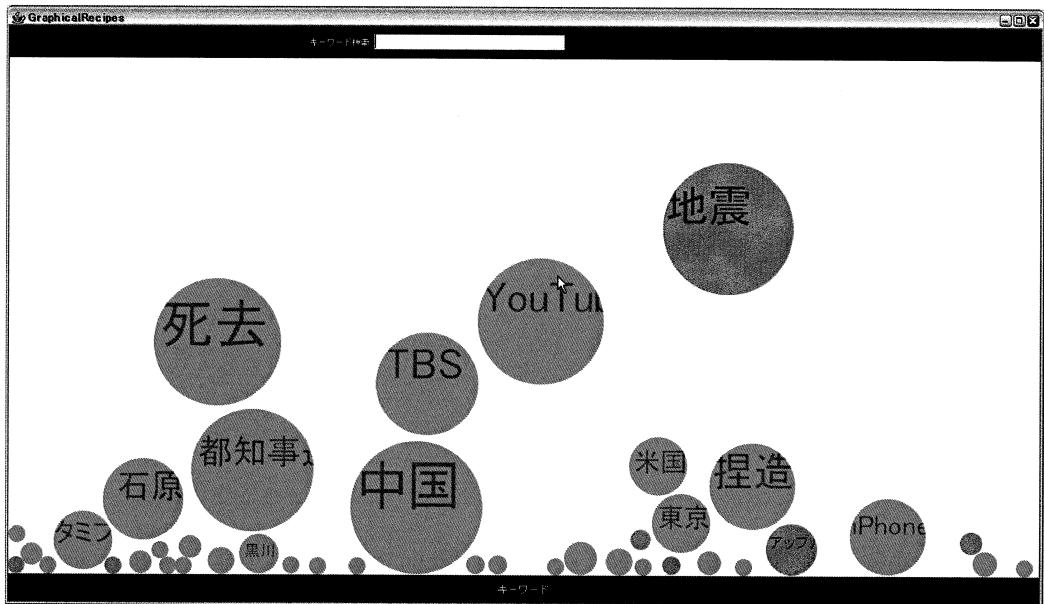


図1 ニュース情報のための試作システムのインターフェイス  
Fig.1 Visualization interface for news.

象としており、本稿で扱うような動的な情報源に対する問題は考慮していない。

### 3. システム

#### 3.1 システム概要

本システムの概要について述べる。本システムでは、ユーザが登録した RSS を定期的に取得し、更新の有無を確認する。更新があった場合には、そのすべての記事からキーワードを抽出する。記事のすべてと抽出したキーワードをデータベースに登録する。

キーワードはキーワードオブジェクトとして視覚化され、画面内に出現する。キーワードオブジェクトの色や大きさによって、各記事におけるキーワードの出現状況を視覚的に得ることができる。

キーワードオブジェクトを選択すると、そのキーワードと記事タイトル内において共起関係のあるキーワードオブジェクトが周囲に表示される。

#### 3.2 キーワードオブジェクト

システムによって抽出されたキーワードは、キーワードオブジェクトとして視覚化されユーザに提示される(図2)。RSS を取得して得た新着記事から抽出されたキーワードは、キーワードオブジェクトとして、画面内に上から落ちてくるように画面内に出現する(図3)。キーワードオブジェクトによるキーワードの視覚化例を図2に示す。

キーワードオブジェクトのサイズは、そのキーワードが記事で使用されてからの経過時間を表している。サイズによって使用されてからの経過時間が分かるようにするため、そのキーワードが新着記事において使用された時点を最大として、時間が経過するに従い、サイズを小さくしていく。あらかじめ設定された時間が経過し、サイズが最小値に達したキーワードは、画面上から削除されていく。既に、一度表示され画面内に表示されている場合には、サイズを最大値まで弾けるように大きく変化させることで、ユーザが一目で再度そのキーワードを用いた記事が発生したことを分かるようにする。

キーワードオブジェクトの色は、設定された一定期間毎のキーワードの出現数を比較した際の増加率（減少率）をあらわしている。増加率がプラスの場合暖色系の色（オレンジ）を用い、増加率がマイナスの場合寒色系の色を用いて表現する。例えば、1日と設定した場合、前日にそのキーワードが出現した数と、当日の数によって算出し、前日と比較しキーワードの出現数が多くなっていれば暖色系の色となり、逆に減少していれば寒色系の色ということになる。

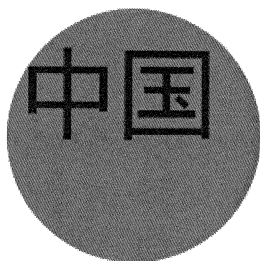
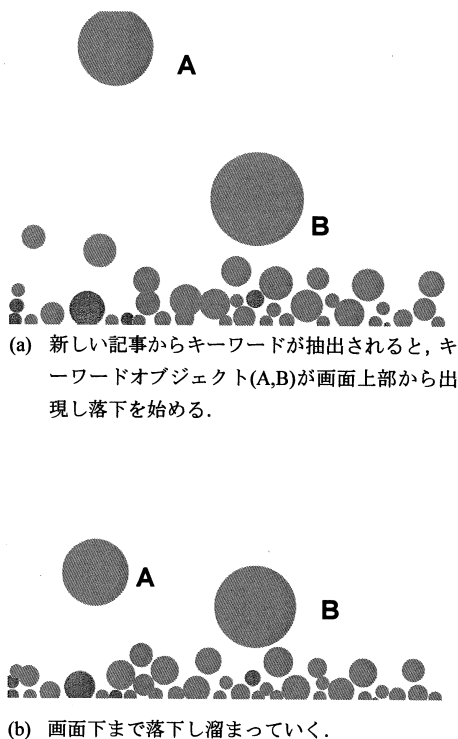


図2 キーワードオブジェクト  
Fig.2 keyword Object.

### 3.3 キーワードオブジェクトの追加

画面上に表示されるキーワードオブジェクトは、RSSによってシステムが取得した記事に基づいて表示される。ユーザは、画面上に表示されていないキーワードを、キーボードから入力するか、もしくは、キーワード一覧画面（図4）の中から選択することによってキーワードオブジェクトを追加することができる。キーワード一覧画面とは、システムによって過去の記事から抽出されたキーワードを選択するための画面であり、画面下のボタンをクリックすることにより、表示と非表示を切り替えることができる。

通常、表示されたキーワードオブジェクトは、再度用いられず一定時間が経過した場合、画面上から削除するが、ユーザによって表示したキーワードオブジェクトは、ユーザによって削除されるまで画面上に提示し続け、削除を行わないこととした。これは、ユーザは、好きなキーワードを用いた記事へ容易にアクセスすることを可能するためと、画面上に表示させておくことで、今後そのキーワードを用いた記事が出た場合に、すぐ見つけられるよう観察対象とすることができるようにするためである。



(a) 新しい記事からキーワードが抽出されると、キーワードオブジェクト(A,B)が画面上部から出現し落下を始める。

(b) 画面下まで落下し溜まっていく。

(c) 時間が経過するにつれオブジェクトのサイズが小さくなっていく。

(d) 既に表示されているキーワード(C)が記事で用いられた場合には、再度膨らんで大きくなる。

図3 キーワードオブジェクト出現時のアニメーションの様子

Fig.3. Animation of keyword object

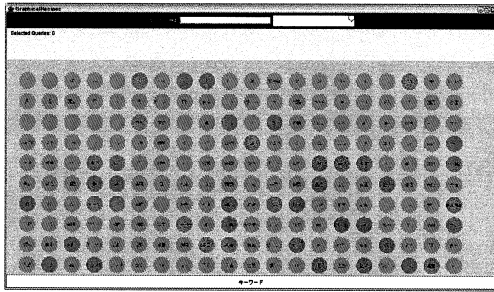


図4 キーワード一覧画面  
Fig.4 keyword list menu.

### 3.4 共起キーワードの提示

キーワードオブジェクトを選択すると、システムが取得した記事において、共起関係のあったほかのキーワードオブジェクトが周辺に表示される(図5)。

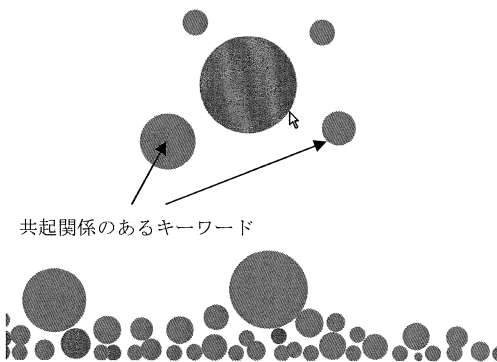


図5 共起関係のあるキーワードの提示  
Fig.5 co-occurrence keywords.

### 3.5 実装

本稿で提案するシステムを、Java 言語 (JDK 5.0) を開発言語とし、データベースエンジン hsqldb (version 1.8.0.7) を利用し、Java アプリケーションとして実装した。RSS 取得には、Informa のライブラリを使用し、形態素解析には、Sen を利用した。

#### 3.5.1 キーワードと共起関係の抽出

キーワードは、RSS で取得した記事データから形態素解析によって自動的に抽出した。

まず、タイトルに使用されている語は、キーワードになる可能性が高いと考え、形態素解析により名詞と未知語を抽出する。その際、単純に名詞を最小単位で切り出し、すべてキーワードとしてしまった場合、キーワードとしては適切でなくなってしまうことが考えられる。例えば、「高校野球」は、「高校」と「野球」に切り分けら

れてしまう。そこで、個別に名詞を切り出すのではなく、連続で出現した場合には、ひとつの名詞として連結してキーワードとして登録する。名詞の連結は、Sen の「連結品詞機能」を用いた。

本実装では、タイトル内におけるキーワードの共起関係のみを抽出しデータベースに記録した。

## 4. シナリオ

我々が想定している本システムの理想的な利用スタイルは、24時間起動しているマシンに、サイトの更新内容用のプログラムを実行し続け、一方で、壁面もしくは大型のディスプレイに継続的に情報を提示し続けるというものである。ディスプレイはタッチディスプレイとなっていて、気になった情報は簡単なインタラクションでその場でアクセスできるというのが望ましい。ユーザは、普段の生活の中で、様々な作業の合間や、作業しながら横目で見ること、更新に気がつくことができ、気になった情報にアクセスすることができる。

## 5. 議論

図1の例は、主に、2007年4月7日から2007年4月15日までの9日間不定期に取得したデータを視覚化した実行例である。地震やタミフルや東京都知事選関連のニュースから抽出されたキーワードが表示されており、この時期に話題になっていたニュースのキーワードが大きく表示された。この視覚化の際、出現数を制限するため、一度しか使われなかったキーワードは省いてしまったので、取得したデータ数が少なすぎたこともあり、表示されないキーワードがあり、アクセスできないニュースも出てしまった。この点に関しては、出現回数が少ないキーワードでも表示しアクセスできるようにする必要がある。

今回の実装では、新着記事においてキーワードが出現した際、該当するキーワードオブジェクトのサイズを大きくすることで、視覚的に目立つように視覚化を行った。これにより、全体の状況を知る目的においては、どのようなキーワードが確認時において使われているのかが分かるために視覚化手法は適切だったと考える。しかしながら、提案するインターフェイスを、記事を読むためのツールとして捉えた場合、ユーザにとっては、未読記事があるかどうか興味の対象となるが、円のサイズの連続的な変化の段階は瞬時に読み取ることが難しいために、未読記事があるかどうかの判断ができない場合が何度かあった。このことから、ユーザの記事の購読状況、つまり、1)未読記事あり、2)未読記事なし、の上記2状態を判別できるようにするため、視覚的に明確に差別化する必要があることが分かった。代替案としては、1)サイズを2種に絞る、2)未読の場合には円の外周にさらに円をつけるといった形態的な変化をつける、などの改善が考えら

れる。

今回の実装では、キーワードとして、タイトルを形態素解析して得られた名詞と未知語、また連続する名詞のみを選定した。そのため、一般的すぎる名詞などのキーワードとしては必ずしも適していない語句がキーワードとして登録されてしまい、結果として多数カウントされたために、視覚的に一番目立ってしまう場合があった。一方で、逆に抽出できなかったキーワードや、表記の揺れによって同一の事柄でも別なキーワードとして登録されることもあった。

キーワード抽出の精度の問題や、表記の揺れの問題を改善するためには、辞書を利用する方法や、キーワード抽出するためのアルゴリズム[5]などと併用し実装することで改善が見込まれると考えている。

## 6. おわりに

本稿では、動的に発生するニュース情報のための視覚化インターフェイスを提案した。

運用期間が短かったため、十分なデータが取得できなかったため、より長期的に利用した場合の結果についても調べる必要がある。今後は、より長期的な運用結果に基づいて、システムおよびインターフェイスの改善をしていきたいと考えている。また、本稿では、個々の記事へのアクセスする部分のインタラクションや、そのための視覚化についてはほとんど考慮しなかった。ツールとして利用していく場合には、個々の記事を読む部分を考慮した改善が必要になってくると考えられる。

### 参考文献

- [1] 石原, 三末, 田中: 木構造の波紋表現 一時系列とカテゴリの同時表現手法—, インタラクション 2006, pp.109-116.(2006)
- [2] Ricardo A. Baeza-Yates, Berthier Ribeiro-Neto.: Modern Information Retrieval, Addison-Wesley Longman Publishing Co., Inc., Boston, MA. (1999)
- [3] Williamson, C., Shneiderman, B.: The dynamic HomeFinder: evaluating dynamic queries in a real-estate information exploration system, Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval, pp.338-346, June 21-24,.(1992)
- [4] Schneiderman, B.: Dynamic Queries for Visual Information Seeking; IEEE Software, v.11 n.6, pp.70-77. (1994)
- [5] 松尾 豊、石塚 満: 語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム、人工知能学会誌、Vol.17、No.3、pp.213-227, 2002