

付箋アノテーションを用いた情報共有システムの試作

Implementing an Information Sharing System using Sticky-Note Annotations

佐野 博之 大園 忠親 新谷 虎松
Hiroyuki Sano Tadachika Ozono Toramatsu Shintani

名古屋工業大学大学院 工学研究科 情報工学専攻

Dept. of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology

We implemented a sticky-note annotation system for web pages. The system allows users to put annotations on contents in web pages. The agent analyses contents attached by annotations, and makes bidirectional links between annotations which are close in meanings. By using the links, users can check other users' annotations and know contents which other users pay attention to.

1. はじめに

Web ページのマルチピック化が進んでいる。ひとつの Web ページ中には、様々なコンテンツが含まれる。Web ページ中の一部を指定するための機能が求められている。

人が書物を参照する際、何度も閲覧するであろう頁に対して、付箋を貼付けるといった動作を行う。WWW 上の情報を参照する際には、Web ブラウザのブックマーク機能を用いて、何度も閲覧するであろう Web ページをブックマークに登録する。書物に対して貼付ける付箋は、貼付ける場所によって頁内の特定の場所を指定することが可能であるのに対し、既存のブックマーク機能は、Web ページのタイトルと URL のみを保存しておくだけのものであり、Web ページ内の特定の場所を指定することを目的としていない。

我々は、付箋を用いて Web ページ中のコンテンツを特定可能とするシステムを提案した [佐野 07]。このシステムでは、貼付け対象となるコンテンツの DOM 要素の子要素として、付箋の画像のタグを付加する。したがって、ユーザの Web 閲覧環境が変化しても、もともと指し示していたコンテンツから付箋がずれることはない。

書物に対して付箋を貼付ける場合、その付箋に対してメモ書きを行うことがある。それと同様に、本システムでは、貼付けた付箋に対してユーザが自由にコメントを付けられる機能を用意する。これにより、ユーザが Web ページ上のコンテンツに対して付箋を貼付け、その付箋を用いて、コンテンツに対してアノテーションを与えることを可能にする。付箋という形でアノテーションを可視化することを、本論文では付箋アノテーションと呼ぶ。ユーザは、Web ページ内に含まれるテキストデータや画像などに対して、付箋によるアノテーションを行うことができる。

本システムでは、ユーザの Web 閲覧環境の変化に頑健な付箋の表示が可能である。ブラウザのレンダリング結果によらず、貼付けた付箋は、もともと指し示していたコンテンツからずれることがない。付箋により、コンテンツを特定可能である。付箋を他のユーザと共有することで、他のユーザが着目しているコンテンツを知ることができる。本論文では、ユーザが貼付けた付箋の応用として、付箋を多数のユーザで共有するシ

ステムについて述べる。他のユーザが貼付けた付箋を確認できるようにすることで、自分以外の他のユーザがこういったコンテンツに着目しているかを知ることができる。本システムでは、関連が高いと思われるコンテンツに対して貼付けられた付箋同士に双方向リンクを作成し、他のユーザが貼付けた付箋を閲覧できるようにした。

ただし本論文では、ユーザは無意味な付箋を貼付けることはしないとする。通常、書物に対して貼付ける付箋は、読者が重要であると感じた箇所に対して貼付けられる。付箋を貼付けるという行為は、情報閲覧の効率を高めるためのものである。重要である箇所と重要でない箇所に対して貼付けられた付箋が入り交じると、情報閲覧の効率は改善されない。したがって、本システムの使用に関しても、ユーザが重要なコンテンツに対してのみ付箋を貼付けるという行為に焦点を当てる。

2. Web ページへのアノテーション

Web ページに対してアノテーションを付加するためのシステムは、すでに多数存在する。有名なものとして、Annotea^{*1} [Koivunen 05] [Kahan 02] が存在する。Annotea とは、W3C^{*2} で開発されている、Web でアノテーション機能を実現するためのフレームワークである。ユーザのアノテーションを活用し新たなサービスに応用しようとする点で、Annotea と本システムの目的は一致している。アノテーションを新たなサービスに応用するためには、ユーザから多くのアノテーションを獲得する必要がある。Annotea を利用するためには、ユーザは、専用の拡張機能をインストールした Firefox^{*3} や、W3C が開発したオープンソースの Web ブラウザである Amaya^{*4} を使用する必要がある。これらのブラウザを利用していないユーザは、Annotea を利用するために新たなブラウザを導入する必要があるため、多くのユーザからアノテーションを獲得することが困難である。本システムは、ユーザが使い慣れた既存の Web ブラウザに対して、プロキシサーバを指定するだけで使用可能である。また、付箋という見慣れた形でアノテーションを行うことが可能であるため、ユーザからの積極的なアノテーションが期待できる。

連絡先: 佐野博之, 名古屋工業大学大学院 工学研究科 情報工学専攻, 〒466-8555 愛知県名古屋市昭和区御器所町, Tel:052-733-6550, Fax:052-735-5584, E-Mail:hsano@toralab.ics.nitech.ac.jp

*1 <http://www.w3.org/2001/Annotea/>

*2 <http://www.w3.org/>

*3 <http://www.mozilla.com/firefox/>

*4 <http://www.w3c.org/Amaya/>

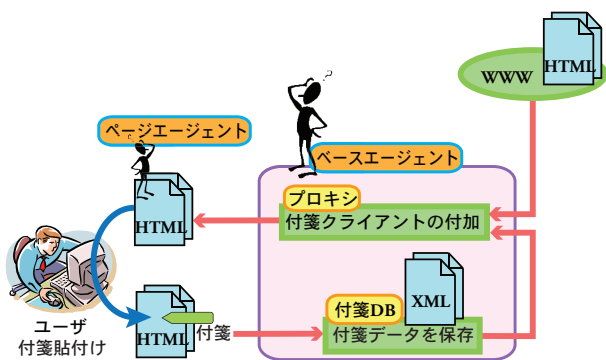


図 1: システム構成図

3. 付箋アノテーションシステム

3.1 システムの構成

図 1 は、本システムの構成図である。本システムは、MiSpider[深菅 06] と呼ばれる Web エージェントモデルに基づき、ページエージェントとベースエージェントから構成される。ページエージェントは、ユーザと直接インタラクションが可能なエージェントであり、Web ブラウザ上で動作する。ページエージェントは、ベースエージェントのセンサ・アクチュエータとして動作する。ベースエージェントは、サーバ側で動作するエージェントである。ベースエージェントは、ページエージェントから得られた情報を基に処理を行う。

Web ブラウザ上で動作する付箋クライアントは、本システムではページエージェントとして動作する。付箋クライアントはユーザに対して付箋を貼付けるためのインターフェイスを提供するシステムであり、JavaScript で実装されている。

ユーザが本システムに Web ページ取得の要求を行うと、Web ブラウザは、プロキシサーバを通じて Web ページを取得する。プロキシサーバは、Web ページを Web サーバから取得し、ページエージェント(付箋クライアント)を付加してから、Web ブラウザに送信する。その際、付箋の情報が保存されているデータベースにアクセスし、過去にその Web ページに対して付箋が貼付けられたかどうかを確認する。付箋が貼付けられていた場合は、データベースから付箋を抽出し、Web ページに付箋を付加する。これらプロキシサーバと付箋データベースは、ベースエージェントとして動作する。

ユーザが新たに付箋の貼付けを行ったり、既存の付箋の更新を行った場合には、ページエージェントが付箋の情報をベースエージェントに送信する。ベースエージェントは、付箋の分類を行い、関連が高いと思われる付箋同士に双方向リンクの作成を行う。

3.2 実行例

実際に本システムを用いて Web ページ内部に付箋を貼付けた時のスクリーンショットを、図 2 に示す。図 2 は、Apple 社の Web ブラウザである Safari*5 を使用して、Google ニュース*6 を開き、ニュース記事 2 つに対して付箋を貼付けた画面である。

ユーザは、図 2 のように、コンテンツに対して付箋を貼付けることができる。貼付けた付箋をダブルクリックすることにより、その付箋情報の詳細を、ウィンドウ(付箋詳細ウィンドウ)



図 2: 実行例

ウと呼ぶ)で閲覧可能である。この実行例では、2 枚の付箋を貼付けている。ここでは、図 2 上部の付箋の付箋詳細ウィンドウが表示されている。次に、図 2 の付箋詳細ウィンドウに示した (1) ~ (6) の部分について説明する。(1) は付箋の貼付け日時の表示である。(2) は付箋の色を選択するために使用する。(3) はコンテンツに対する注釈であり、これをアノテーションとして用いることが可能である。ここに入力された注釈は、付箋の画像の上にも重ねて表示される。(4) は関連する付箋へのリンク表示であり、表示されている文字列をクリックすることにより、付箋が貼られたコンテンツと関連しているコンテンツに対して貼付けられた付箋を開くことができる。このリンクに関しては、次章で詳しく述べる。なお、(5) は付箋を削除するためのボタンであり、(6) は付箋詳細ウィンドウを閉じるためのボタンである。

4. 双方向リンクに基づく付箋の共有

本章では、双方向リンクに基づく付箋の共有について述べる。ブログにはトラックバックという機能が存在する。トラックバックとは、リンク先のサイトに、リンク元のサイトの情報を通知することができる機能である。トラックバックによって、ブログに双方向性が生まれる。本システムでは、トラックバックの考え方にに基づき、関連するコンテンツ間を相互にトラバース可能なリンクを自動生成する。本論文では、このリンクを双方向リンクと呼ぶ。

4.1 biLink エージェント

本システムでは、biLink エージェントと呼ばれる Web エージェントが、ユーザの付箋貼付けを監視している。biLink エージェントは、付箋に対して双方向リンクを与える機能を持つ。ユーザが付箋を貼付けると、ページエージェントは、DOM ツリーに基づき、付箋が貼付けられたコンテンツを抽出する[伊藤 07]。抽出されたコンテンツは、ベースエージェントに送信される。ベースエージェントは、受信したコンテンツの内容を解析し、付箋の分類を行う。付箋の分類手法に関しては、4.2 節で詳しく述べる。そして、関連性が高いと思われるコンテンツに貼られた付箋同士に、自動的に双方向リンクを作成する。

図 3 は、biLink エージェントの動作の例である。この例では、ユーザが本システムを用いて、(1) ~ (6) の計 6 枚の付箋を貼付けたとする。(A) は全てのユーザが貼付けた付箋の全体

*5 <http://www.apple.com/safari/>

*6 <http://news.google.co.jp/>

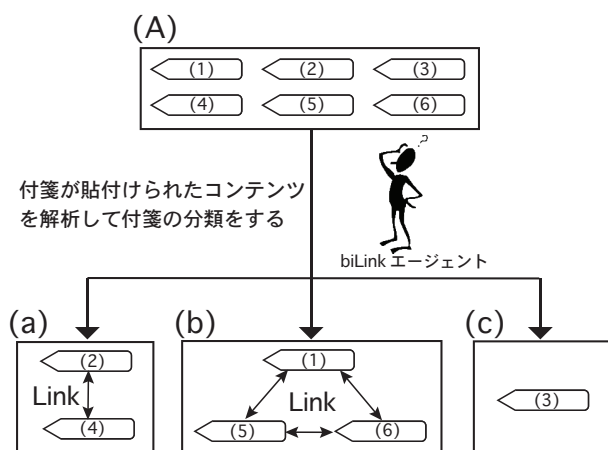


図 3: biLink エージェントの動作例

集合であり、biLink エージェントは、これらの付箋の分類を行う。図 3 で、矢印は付箋の分類を表す。これらの付箋が、以下のように (a), (b), (c) の 3 つに分類されたとする。

- (a) = {(2), (4)}
- (b) = {(1), (5), (6)}
- (c) = {(3)}

biLink エージェントは、この分類の結果をもとに、付箋同士に双方向リンクを作成する。今、(a) には (2) と (4) の付箋が存在するため、これら 2 つの付箋を双方向リンクで結ぶ。(b) には (1) と (5) と (6) の付箋が存在するため、これら 3 つの付箋を双方向リンクで結ぶ。(c) には (3) の 1 つの付箋しか存在しないため、リンクを作成する必要はない。

ここで、(2) の付箋がユーザ A によって貼付けられ、(4) の付箋はユーザ B によって貼付けられたものであった場合、ユーザ A とユーザ B は、双方向リンクを通じて、お互いの付箋を閲覧可能となる。ユーザ A は、自分が貼付けた付箋 (2) に対して作成されたリンクをクリックすることにより、ユーザ B によって貼付けられた付箋 (4) を閲覧可能である。同様に、ユーザ B も、ユーザ A によって貼付けられた付箋 (2) を閲覧可能である。このように、biLink エージェントによって作成された双方向リンクを辿り、他のユーザが貼付けた付箋を確認することで、他のユーザが着目しているコンテンツを知ることが可能となる。

4.2 付箋分類手法

biLink エージェントが付箋の分類を行う手法について述べる。biLink エージェントは、付箋が貼付けられたコンテンツの存在する Web ページを、MeCab^{*7}を用いて形態素解析し、文書の索引語を決定する。今回は実装を簡単にするために、索引語の評価値として tf-idf の値を用いた。tf を求める際のベースとなる文書は、付箋が貼付けられたコンテンツの存在する Web ページである。本システムは Web をベースとしたアプリケーションであるため、idf を求める際の全文書数として、Yahoo!API^{*8}で検索可能な全ての Web ページの数を用いた。idf を求める際の索引語の登場する文書数として、その索引語の Yahoo!API での検索結果の数を用いた。

*7 <http://mecab.sourceforge.net/>

*8 <http://developer.yahoo.co.jp/>

付箋の分類を行う際の文書の類似度は、ベクトル空間モデルに基づき、2 つの文書ベクトルのコサイン尺度から求める。文書ベクトルは、各次元に索引語を割当て、各成分に索引語の評価値を割り当てたものとするが、付箋が貼付けられたコンテンツに重みをかけて計算する。即ち、本システムにおける文書ベクトルの計算式は、次のようになる。

$$d_i = (w_{i1}, w_{i2}, \dots, w_{iM})^T + \alpha(v_{i1}, v_{i2}, \dots, v_{iM})^T$$

ここで、 w_{ij} は i 番目の文書における語 t_j の評価値 ($j = 1, 2, \dots, M$)、 v_{ij} は i 番目の文書の付箋を貼ったコンテンツにおける語 t_j の評価値、 M は文書集合に含まれる異なる語の数、 α は重みである。

次に、2 つのベクトルを d_1, d_2 とした場合、コサイン尺度は次の式で表せる。

$$\cos(d_1, d_2) = \frac{d_1 \cdot d_2}{\|d_1\| \|d_2\|}$$

コサイン尺度は、2 つのベクトルの角度が小さい場合に、大きな値を取る。また、付箋を分類したクラスタについては、クラスタに含まれる付箋の文章のベクトルの平均ベクトルを、クラスタのベクトルとする。

biLink エージェントが付箋の分類を行うために用いる手続き *ClassifyStickers* を、図 4 に示す。*ClassifyStickers* では、ユーザが初めて付箋を貼付けたコンテンツを解析し、文書ベクトルを求め、それを 1 つのクラスタとする。以降、新しく付箋が貼付けられたコンテンツを解析し、文書ベクトルを求め、その文書ベクトルと、既存のクラスタの文書ベクトルとのコサイン尺度を計算する。コサイン尺度が閾値よりも大きければ、付箋をそのクラスタに追加し、クラスタのベクトル更新をする。コサイン尺度が閾値を下回っていれば、新しくクラスタを作成し、付箋をそのクラスタに分類する。

ユーザが本システムの使用を開始した時点では、付箋の数は 0 であるため、クラスタ数も 0 である。ユーザが付箋を 1 つ貼付けた場合に、初めてクラスタが発生する。その後、付箋が増えるたびに、biLink エージェントはクラスタリングを行い、同じクラスタに属する付箋が発生した場合には、双方向リンクの作成を行う。双方向リンクを作成する際には、付箋をノードとしてグラフで表現した場合、完全グラフとなるように双方向リンクで結ぶ。ただし、HTML でリンクを実現するためのアンカー要素の href 属性では、リンク先を一つしか指定できない。このため、本システムでは、3.2 節で示した図 2 の付箋詳細ウィンドウにおいて、(4) のように全てのリンクを羅列することにより解決を図っている。

5. 考察

ユーザが重要だと思ったコンテンツに対して付箋を貼付けることにより、biLink エージェントが、それらのコンテンツを付箋を通じてリンクで結んでゆく。付箋の分類を行う際に、付箋の貼付け対象となったコンテンツを用いたクラスタリングを行った。コンテンツに含まれるテキストノードのみを用いて文書ベクトルを作成すると、索引語の数が十分に得られず、クラスタリングがうまく行われなかった問題があった。本システムでは、Web ページをベースに文書ベクトルを作成し、付箋を貼付けたコンテンツに対して重みをかけている。そのクラスタリング結果を基に、biLink エージェントは双方向リンクの作成を行う。今後の課題として、クラスタリングの妥当性を検証する必要がある。

```

入力：貼付けた付箋 fusen
      既存のクラスタ集合
       $G = \{g_1, g_2, \dots, g_n\}$ 
      既存のクラスタのベクトル
       $D = \{d_1, d_2, \dots, d_n\}$ 
procedure ClassifyStickies (fusen,  $G$ ,  $D$ )
begin
   $d_f \leftarrow fusen$  の文書ベクトル
  //既存のクラスタ数が 0
  if  $n == 0$  then
    // fusen を一つ目のクラスタとする
     $g_1 \leftarrow \{fusen\}$ 
     $d_1 \leftarrow d_f$ 
     $G \leftarrow \{g_1\}$ 
     $D \leftarrow \{d_1\}$ 
  else
     $max \leftarrow -\infty$ 
     $s \leftarrow 0$ 
    // コサイン尺度が最も大きいクラスタ
    for each  $d_i \in D$  do
      if  $max < \cos(d_i, d_f)$  then
         $max \leftarrow \cos(d_i, d_f)$ 
         $s \leftarrow i$ 
      endif
    enddo
    if  $max > \text{閾値}$  then
      // クラスタに追加, ベクトル更新
       $g_s \leftarrow g_s \cup \{fusen\}$ 
       $d_s \leftarrow g_s$  の平均ベクトル
    else
      // fusen を新たなクラスタとする
       $g_{n+1} \leftarrow \{fusen\}$ 
       $d_{n+1} \leftarrow d_f$ 
       $G \leftarrow G \cup \{g_{n+1}\}$ 
       $D \leftarrow D \cup \{d_{n+1}\}$ 
    endif
  endif
end.

```

図 4: ClassifyStickies

付箋にはコメントを書き込む機能が用意されているため、ユーザが、コンテンツに対して補足説明を行ったり、自分の感想を与えたりなど、メタデータを付加することができる。付箋の共有により、メタデータを複数のユーザ間で共有することが可能となり、情報共有が実現する。応用として、本システムを用いたオンライン議論が考えられる。大学では、講義の内容を Web ページとして学生に公開している場合がある。こういった Web ページに対して本システムを用いることにより、その Web ページに掲示板のような発言をする場が用意されていないとしても、付箋のコメントを用いて授業内容に対して補足的な事項を付加したり、質問やその回答などを Web ページに対して書き込むことが可能となる。付箋の共有により、講師や他の学生が与えたコメントを閲覧することができる。また、関連するコンテンツへのリンクが作成されるため、講義の Web ページ以外へ貼られた付箋は、その講義の参考資料として有用である可能性がある。

また、本システムを用いることは、人気のあるコンテンツの発見を容易にする。WWW のユーザは、情報収集をする際

に検索エンジンを使う。通常の検索エンジンでは、サーチエンジン独自の検索アルゴリズム [兼宗 04] を用いて、ページをランキングしている。ユーザは、検索結果の上位にランクされた Web ページを重要なものとみなし、上位にランクされた Web ページは積極的に開くが、下位にランクされた Web ページを開くことは少ない。たとえ検索結果が下位にランクされる Web ページであっても、本システムでは、ユーザ自身が Web ページの重要だと思ったコンテンツに積極的にアノテーションを行うことにより、その Web ページはより一層価値を増す。貼付けた付箋を多数のユーザ間で共有することで、検索エンジンに頼らない情報収集を行うことが可能である。ただし最初に述べたように、これは、付箋が重要なコンテンツに対して貼付けられると限定した場合である。付箋の利用方法はユーザによって異なり、実際には、重要でないコンテンツに対して付箋を貼付けるユーザが存在する可能性もある。したがって、複数の被験者に本システムを使用させることで、ユーザの付箋の使い方を明らかにする必要がある。

6. おわりに

本論文では、Web ページに存在するコンテンツへの付箋アノテーションシステムの応用として、双方向リンクに基づき、付箋を複数のユーザ間で共有するシステムの試作を行った。付箋画像を DOM ノードに対して貼付けることにより、閲覧環境に頑健な付箋が実現する。ユーザの Web 閲覧環境が変化しても位置がずれない付箋を実現することで、Web エージェントが付箋を貼付けたコンテンツを解析し、内容的に関連が深いと思われる付箋同士に自動的に双方向リンクを作成することが可能となった。双方向リンクにより、情報共有が実現すると考えた。ユーザは双方向リンクを辿ることにより、他のユーザが着目しているコンテンツを閲覧可能である。また、付箋に与えられたコメントから、他のユーザが、そのコンテンツに対してこういった考えを持っているかを知ることができる。

参考文献

- [深菅 06] 深菅裕二郎, 大園忠親, 新谷虎松, “回線切断に頑健な Web エージェント MiSpider について,” 日本ソフトウェア科学会第 23 回大会論文集, Sept. 2006.
- [伊藤 07] 伊藤太樹, 浅見昌平, 大園忠親, 新谷虎松, “携帯電話のためのコンテンツ分割に基づくナビゲーションエージェントの試作,” 合同エージェントワークショップ & シンポジウム 2007 (JAWS 2007), Oct. 2007.
- [Kahan 02] Jos Kahan, Marja-Riitta Koivunen, Eric Prud'Hommeaux, and Ralph R. Swick, “Annotea: An Open RDF Infrastructure for Shared Web Annotations,” 10th International World Wide Web Conference (WWW10), Hong Kong, May. 2001.
- [兼宗 04] 兼宗進, “検索エンジンの検索アルゴリズム,” 情報の科学と技術, Vol. 54, No. 2, pp. 78-83, 2004.
- [Koivunen 05] Marja-Riitta Koivunen, “Annotea and Semantic Web Supported Collaboration,” ESWC 2005, UserSWeb workshop, 2005.
- [佐野 07] 佐野博之, 大園忠親, 新谷虎松, “DOM ツリー解析に基づく Web ページへの付箋貼付けシステム,” 情報科学フォーラム論文集, Sept. 2007.