

電子掲示板における記事情報の効率的活用のための一考察 – microformats を用いて –

小林曜峰, 伊藤一成, Martin J. DÜRST

青山学院大学理工学部情報テクノロジー学科

1 はじめに

電子掲示板は多くの人に利用されている Web システムであるが, その記事構造が単純であるため, 情報の検索, 再利用は容易ではない. 掲示板からの情報取得に対しては自然言語処理を用いる研究が藤村ら [1] や峠ら [2] 等により行われているが, 取得した情報の再利用性は必ずしも高くない. 一方, 一部の掲示板ではメタデータを付加し情報を構造化している. そのような情報は掲示板システム側が提供するサービスでしか利用できないが情報の構造化をオープンフォーマットを用いて行うことができれば, システムにとらわれず, 情報活用の幅を大きく広げることができるであろう.

そこで, Web 構造化手法の一つである microformats¹を利用した掲示板を作成し, 記事情報の構造化と活用の展望について述べる. microformats は半構造化された情報をより簡単に表示, 索引付け, 追跡することができる [3] オープンフォーマットであり, 既存の (X)HTML 環境で利用できる. microformats を用いることにより, 掲示板上的の記事情報をより効率的に扱うことが可能になる.

2 記事構造への microformats の利用

microformats では, それぞれの HTML タグの属性に対し特定の値を設定することで, そのタグに囲まれた範囲を意味づけすることができる. 用途別に複数の仕様が策定されており, それぞれの仕様により Web ページやハイパーリンクのマークアップが可能である.

このような特徴を利用して掲示板の記事内容や記事内のリンク等への意味づけと構造化を行う.

2.1 記事内容のマークアップ

microformats の利用により, イベント, 予定等の記事やレビュー, 評価関連の記事の意味づけが可能である. これらは hCalendar や hReview を用いて記事全体を意味づけ, 構造化することができる. 図 1 及び図 2 に hReview を用いてマークアップした記事の例を示す.

また, hCard を用いれば各記事の投稿者情報のマークアップが可能である. なお, 投稿者情報にはニック

A Discussion about Efficient Use of Article Information in a Bulletin Board System – Using microformats
Terumine KOBAYASHI, Kazunari ITO and Martin J. DÜRST
Department of Integrated Information Technology, College of Science and Engineering, Aoyama Gakuin University
5-10-1 Fuchinobe, Sagamihara, Kanagawa 229-8558, Japan

¹<http://microformats.org/>

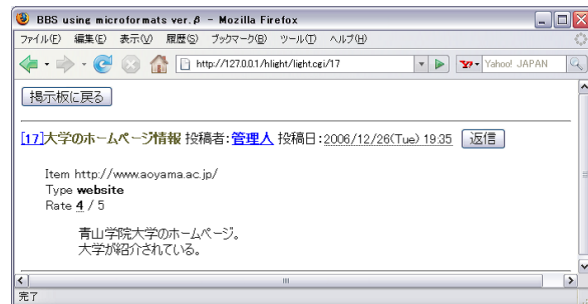


図 1: hReview を用いてマークアップした記事例

```
<div class="hreview">
  <a href="17">[<b>17</b>]</a>
  <span class="summary">大学のホームページ情報</span>
  投稿者: <span class="reviewer vcard">
    <span class="vcard">
      <a href="mailto:kanri@mail.com" class="email fn">
        <span class="fn">管理人</span>
      </a>
    </span>
  </span>
  投稿日:
    <abbr class="dtreviewed" title="20061226T1935">
      2006/12/26(Tue) 19:35
    </abbr>
  Item <span class="item">
    <span class="fn">http://www.aoyama.ac.jp/</span>
  </span><br>
  Type <abbr class="type">website</abbr><br>
  Rate <abbr title="4" class="rating">4</abbr> / 5<br>
  <blockquote class="description">
    青山学院大学のホームページ。
    大学が紹介されている。
  </blockquote>
</div>
```

図 2: マークアップした記事のソース例 (抜粋)

ネームが用いられる場合も多いが, hCard にはニックネームも利用可能である.

2.2 リンクに対するマークアップ

記事内には「掲示板上的の他の記事」や「外部情報」に対するリンクが存在する.

これらの中で, リンクが評価, 意見等に用いられる場合は, VoteLinks を用いて「賛成」「反対」あるいは「どちらでもない」といったメタデータを付加できる. また, リンク先が人物を示す場合, リンク先と記事投稿者との関係を XFN (Xhtml Friends Network) を用

いて記述できる。

外部情報の一例として映像や圧縮ファイル等のコンテンツが挙げられる。このようなリンクに対しては、rel-enclosure を利用することで他のリンクとの差別化が可能になる。

2.3 メタデータの入力

記事の意味づけに利用されるメタデータは多くの場合フォーム等を利用して入力される。複数のメタデータを入力する際にはフォームも複数必要となるため、上記のように多量のメタデータの入力に対しては、ユーザの負担を極力減らすために入力をサポートする必要がある。

このような観点から、作成したシステムでは入力フォームの集約とメタデータの入力サポートを行っている。例えば、hCalendar や hReview のような排他的な選択に対してはその選択に応じたフォームのみを表示する。リンクのマークアップはリンクを半自動的に抜き出すことにより、付加をサポートする。

このような手法を用いることにより、ユーザのメタデータの入力に対するコストを下げるができる。

2.4 メタデータの出力

メタデータは可能な限りユーザに意識させずに表示することが望ましい。そのため記事内の情報を利用し埋め込みを行う。例えば hReview の “reviewer” (記述者) や “dtreviewed” (記述日時) といった要素は記事の投稿者や投稿日と同義であるため、それらの項目内に埋め込める。なお、この方法はメタデータの入力数を減らすという観点からも有益である。

このようにして埋め込めないメタデータのみ、記事と分けて表示させる。

3 記事情報の効率的活用

記事情報を microformats でマークアップするメリットとして主に以下の項目が挙げられる。

3.1 検索手段の多様化

従来、掲示板からの情報検索は主にパターンマッチングによる文字列検索に限られていた。メタデータを付加すれば、それに基づいた検索も可能になる。microformats はオープンフォーマットであるので、利用可能な検索方法は掲示板システムが提供するメタデータ検索だけではない。実際、Technorati 社では microformats 検索サービス²を提供しており、記事情報の URL を登録すれば同サービスを利用した記事検索も可能である。

3.2 情報の再利用性向上

microformats で記述された情報は対応ソフトウェアを利用することで他のアプリケーションで容易に再利用することができる。これにより記事情報を様々な用途に活用することが可能である。一例として、Operator³ という Firefox の拡張機能を利用すればページ上の hCard、hReview、hCalendar 等で記述された情報を抜き出し、

google Calendar⁴ など様々な Web アプリケーション等に転用できる (図 3 参照)。microformats に対応するアプリケーションは多数存在するため、様々な方法で情報を活用することが可能となり、記事情報の再利用性を大きく向上させることができる。



図 3: Operator を用いた投稿者情報の再利用例

3.3 リンクの活用

マークアップされたリンクは、集計や分析により利用価値が高まる。例えば、rel-enclosure で記述されたリンクのみを抜き出せばマルチメディアコンテンツの検索を容易に行うことが可能になるし、VoteLinks の情報を分析すれば、特定の記事に対する他の記事による評価を集計することができる。このような機能を既存の掲示板で実現するのは容易ではないが、microformats を用いれば比較的容易に実現可能である。

4 まとめと今後の課題

掲示板は複数のユーザによって情報が投稿されるため、記事情報も多種多様であり構造化による恩恵は大きい。本手法を用いて掲示板の構造化を行えば、既存の掲示板システムでは難しい、記事情報の効率的な活用が可能になる。

今後は、インタフェースの改良によるメタデータの入出力方法の更なる改善、任意の情報に対する意味づけ手法等を検討していく。

謝辞 本研究の一部は科学研究費補助金 (課題番号: 18700095) の支援によるものです。ここに記して謝意を表します。

参考文献

- [1] 藤村滋, 豊田正史, 喜連川優: 文の構造を考慮した評判抽出手法 (掲示板), 第 16 回データ工学ワークショップ (DEWS2005) 論文集, 6C-i8 (2005).
- [2] 峠泰成, 大橋一輝, 山本和英: ドメイン特徴語の自動取得による Web 掲示板からの意見文抽出, 第 11 回言語処理学会年次大会, pp. 672-675 (2005).
- [3] Khare, R. and Çelik, T.: Microformats: a pragmatic path to the semantic web, *Proceedings of the 15th International Conference on World Wide Web*, pp. 865-866 (2006).

²<http://kitchen.technorati.com/search/>

³<https://addons.mozilla.org/firefox/4106/>

⁴<http://www.google.com/calendar/>