

# キーワードによる文章間のつながりを活用した コミュニティ内情報共有

芝崎 亮<sup>†</sup> 千葉 大作<sup>†</sup> 中沢 実<sup>‡</sup> 服部 進実<sup>‡</sup>

<sup>†</sup>株式会社 アルファシステムズ      <sup>‡</sup>金沢工業大学 知的創造システム専攻

## 1 はじめに

ある特定の目的の為に形成されたコミュニティの参加メンバーに対して、Blog を通じてコミュニケーションを行う試みが登場している。東京虎ノ門にある社会人大学院の金沢工業大学知的創造システム専攻でも、選択科目の受講生に対して Blog を提供し、トラックバックやコメントを利用した教員・受講生間や受講生同士のコミュニケーションを試みている。

しかし、現実にはコミュニティの参加メンバーの大多数が、そのコミュニティの中心人物が運営する Blog を主に閲覧する傾向があり、コメントやトラックバックも中心人物に対してのみ行いがちである。実際に金沢工業大学での授業における Blog の利用でも、授業における中心人物であるといえる教員とのやりとりが大多数を占めた。

そこで本論文では、あるメンバーが投稿した Blog エントリーに対して、関連する Blog エントリーを紹介するシステムを構築する事で、コミュニティ内情報共有を活性化することを目標とした。

## 2 Blog エントリー共有システム

### 2.1 Blog エントリー共有システムの概要

図 1 は単一の Blog エントリーを表示した画面である。コミュニティ内にある全ての Blog には、各エントリーごとにその内容を表すキーワードが複数個選ばれており、エントリーの本文の上に表示されている。Blog 閲覧者はこれらのキーワードの中から一つ選択する事で、全てのエントリーの中から選択されたキーワードを持つエントリーが一覧表示された画面を閲覧する事ができる。図 1 ではキーワードとして『Ping』、『MySQL』、『PHP』、『実装』、『保存』が表示されており、『PHP』を選択し

information sharing among community using the link between document by keyword

Ryo SHIBASAKI<sup>†</sup>, Daisaku CHIBA<sup>†</sup>, Minoru NAKAZAWA<sup>‡</sup> and Shinmi HATTORI<sup>‡</sup>

<sup>†</sup>ALPHA SYSTEMS Inc.

<sup>‡</sup>Kanazawa Institute of Technology

{shibasr, chibad}@alpha.co.jp,

{nakazawa,

hattori}@infor.kanazawa-it.ac.jp



図 1: Blog エントリーの関連キーワード

た場合に表示される画面が図 2 である。



図 2: 同一キーワードを持つエントリーの一覧表示

図 2 の画面では、直前に閲覧していたエントリーも含めて 4 件の『PHP』に関するエントリーを選択できる。さらに、選択されたエントリーに表示されたキーワードを選択する事で、閲覧の対象を『PHP』をテーマにするエントリーから『PHP』に関連するキーワードをテーマにするエントリーへ移動できる。これによって、自分の興味があるテーマのエントリーはもちろんのこと、関連するテーマのエントリーも気軽に閲覧が可能になる。

## 2.2 Blog エントリ共有システムの実装

Blog エントリ共有システムは、コミュニティ内の Blog 群に対して以下の手順を踏む事で実現される。まず、エントリに対して新規作成や修正が行われると、Blog システムは 2.3 キーワード付加で述べる手法を用いて、エントリに付加するキーワードを最大五個まで設定する。そして、投稿されたエントリのキーワードを、各キーワードが設定されたエントリを一覧表示するページへのリンクと一緒に表示する。さらに、Blog システムが自動生成した RSS 1.0 に、オブジェクトの分類を可能にする `mod_taxonomy`[1] を利用したエントリのキーワード情報を付加して、更新情報として公開した。

エントリを一覧表示するページは PukiWiki を利用して構築した。設定された時間が経過された状態で PukiWiki が閲覧されると、PukiWiki は登録された RSS をパースして前回にパースした時から新規作成や修正が行われたエントリを探す。さらに、見つかったエントリのキーワード情報を取得し、各キーワードごとに一覧表示を行うための PukiWiki のページに対してそのエントリのタイトルやリンク、投稿者の情報を追加する。以上によって Blog 同士が同じキーワードでつながることが可能になる。

## 2.3 キーワード付加

Blog エントリに対するキーワードの設定であるが、エントリの文章から特徴語抽出を行い、特徴キーワードとしてエントリに付加することにした。特徴語抽出を行う文章に対して分かち書きを行い、各単語に対して式 (1) で求められる  $tf \cdot idf$  値 [2] を計算し、値が大きい単語を特徴語として選択した。

$$tf \cdot idf = n_i \cdot \log \left( \frac{|D|}{|(d_i \supset t_i)|} \right) \quad (1)$$

$n_i$  : 抽出を行う文章に単語  $t_i$  が登場した回数

$|D|$  : 文章集合  $D$  に含まれる文章の数

$|(d_i \supset t_i)|$  : 文章集合  $D$  に単語  $t_i$  が登場した回数

各項の意味は  $tf$  の項が文章中によく登場する単語が重要であり、 $idf$  の項は他の文章集合にはあまり登場しない単語が重要であることを意味している。文章集合  $D$  には、以下の三種類の文章集合を用意し、システム管理者が変更できるようにした。

- インターネット上の Blog エントリ (100,000 件)
- 授業で受講生が記述した Blog エントリ (96 件)
- IT 分野の用語辞典の説明文 (8669 項目)

## 3 試用対象

金沢工業大学知的創造システム専攻で行われた「WEB アプリケーション開発特論」において、10 名の受講生が Blog に書き込みを行っていた。全 15 回の講義において、各受講生が書き込みを行ったエントリ数は最大が 15 件で、最小が 2 件で、平均が 9.6 件であった。また、一エントリあたりの文字数については、最大が 969 文字、最小が 7 文字、平均が 127 文字であった。これらに対して、本システムの適用を予定している。

## 4 類似サービス

類似のサービスとして、付加されたキーワードを通してブックマークや画像を共有するサービス [3][4] が存在する。これらのサービスと今回のシステムの間には、共有するコンテンツがインターネット上か特定のコミュニティ上かという違いがある。また、類似のサービスはキーワードを手で付加するのに対し、今回のシステムは自動的に付加している。これは、社会人でもある受講生の負担を軽減すると同時に人手による付加により生じるキーワードのゆれを解決する為に行った。キーワードのゆれの問題はコンテンツ量の増加によっても解決されるが [5]、今回の状況では十分な量が得られていなかった。

## 5 おわりに

今後は、実際に受講生の Blog エントリを対象にシステムを適用して、キーワードとエントリのつながりの様子を分析する予定である。三種類の文章集合による特徴語抽出の違いを参考にして、今回のシステムに相応しい特徴語抽出を考案したいと考えている。

## 参考文献

- [1] RSS 1.0 Modules: Taxonomy. <http://web.resource.org/rss/1.0/modules/taxonomy/>.
- [2] Tf-idf. <http://en.wikipedia.org/wiki/Tf-idf>.
- [3] del.icio.us. <http://del.icio.us/>.
- [4] Flickr. <http://www.flickr.com/>.
- [5] Adam Mathes. Folksonomies - Cooperative Classification and Communication Through Shared Metadata. <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>, Dec 2004.