

小グループ向け文書情報共有システムの実装と評価

今井 良美 †

角田 博保 ‡

赤池 英夫 ‡

電気通信大学情報工学専攻†

電気通信大学情報・通信工学専攻‡

1 はじめに

本研究では、小グループにおける文書情報共有システムの実装を行った。

Web 上には多様な文書情報が存在し、様々な検索方法が存在する。特に、研究室においては様々な種類の文書情報が存在し、それぞれに管理されていることが多い。多元管理されていることで、それぞれの関連性を保持できないという問題が発生する。例えば、ゼミの資料とそのゼミに関連するメモなど、関連性は保持しにくいことが多い。

そこで本研究では、文書情報に対して半自動的にタグ付けをし、一元管理するシステムを作成した。ここでは文書情報として、グループ内でのコミュニケーションを支援するシステム Urban[1](後述) で蓄えられたテキスト情報を取り扱う。

2 Urban

Urban(図1)はグループ内での情報源となりコミュニケーションを支援することを目的とし、2003年に本研究室で開発されたコミュニケーションツールである。Urbanのシステムはベースシステムとその上



図 1: Urban の外観

で実行されるモジュールの二つから構成されている。そのため、モジュールの追加変更など環境にあった対応を容易に行うことができる。

モジュールには以下のような機能があり、それぞれのデータはデータベースに保存されている。

● 掲示板機能

通常の掲示板と同様にスレッド方式をとっている。複数のスレッドが作成あるいは更新時刻順に並び、個々のスレッドに対して返信ができるようになっている。

● メモ機能

公開メモ、限定メモ、個人メモの3種類が存在する。公開メモは研究室のメンバー全員が見ることができ、個人メモは作成者しか見ることができない。限定メモは、指定されたメンバーのみがメモを閲覧することができる。

● ゼミ機能

発表者、論文名、発表日など、ゼミに関する情報が存在する。また、当該論文へのリンクが存在する。

● 目標設定機能

各週に対して、先週の成果、今週の目標などのテキスト情報が存在する。それに対しての、コメントが存在する。

3 本システム

3.1 取り扱う情報

掲示板の一記事や一編の論文など、それ以上細かく分割しない最小のまとまりをエンティティと呼ぶこととする。

それぞれのエンティティにはタグが付けられる。タグはエンティティの特徴語および日付や著者が自動で付けられ、一つのエンティティに対し、複数のタグがつく。

タグには属性が付けられる。Kerstinらの分類スキームの研究[2]ではタグを、Topic, Time, Location, Type, Author/Ownerの5種類に分類している。本システムでは、文書情報のみを対象としたためこれをTime, Type, Author/Ownerの3つに限定し、これに特徴語(keyphrase)とユーザ定義語(user)を加えた5つの属性を設定した。一つのタグに対し、必ずこれらの属性を持つようにした。

Urbanへの登録時にエンティティの著者、日付や特徴語を自動で抽出し、これらをタグとして保存する。また、既にUrban上に存在するエンティティに対し、タグの追加、編集、削除を行うことができる。

†Yoshimi IMAI, Department of Computer Science, The University of Electro-Communications

‡Hiroyasu KAKUDA and Hideo AKAIKE, Department of Communication Engineering and Informatics, Graduate School of Informatics and Engineering, The University of Electro-Communications

3.2 システムが扱うデータ

Urban は本研究室で8年にわたる利用実績があり、テキストのみならず、画像、動画など各種メディアのデータが多数蓄積されている。今回、本システムはUrbanのモジュールで扱われるテキストデータをエンティティとして利用する。

3.3 本システムの機能

3.3.1 取り扱う情報の加工

Urban へのエンティティの登録時に、エンティティから特徴語を自動で抜き出し、タグとして保存する。

特徴語の抽出は、YahooAPI[3]のキーフレーズ抽出を使用している。

Urban は Web 上で動くシステムなので、各モジュールのテキスト情報に HTML タグが混在している。そのためまず HTML タグを除外し、特徴語の抽出を行っている。論文に関しては、全文ではなく概要やはじめにの部分を抜き出し、特徴語の抽出を行った。これは論文であれば、この部分に重要なことが書かれていることが多いためである。

また、YahooAPI では特徴語にスコアが付けられ、本システムではスコアが一定以上のものを特徴語として採用した。

これらの特徴語及び、日付データなどから、エンティティに自動的にタグをつける。エンティティに付加するメタ情報は独自に定義した XML アプリケーションとして保存している。タグは、作成日 (time)、エンティティの存在するモジュール (type)、作成者 (author)、特徴語 (keyphrase)、ユーザの定義語 (user) の属性を持つ。フォークソノミーを実現するよう、ユーザがタグの追加、編集、削除することを許容し、そのタグにはユーザの定義語の属性が付けられる。

3.3.2 検索機能

あるタグを対象にして検索すると、そのタグがついたエンティティの集合が得られ、検索結果として表示する。

同時に、そのタグに類似したタグがついたエンティティの集合も求め、これをサブ検索と呼ぶ。これも検索結果として表示する。ただし、本来の検索結果よりも目立たないようにしている。

さらに、これらの検索結果からタグを使い、絞り込み検索をしていくことができる。類似タグによるサブ検索を行うことで、単なるタグによる絞り込みだけでなく、関連のあるエンティティやメタ情報のサジェストが行える。

3.3.3 可視化機能

検索およびサブ検索によって得られたエンティティの集合を、タグクラウドやネットワーク構造として可視化する。

その集合に含まれるエンティティについているタグの出現頻度からタグクラウドを表示したり (図2)、同一タグがついているエンティティを上位にすることなどが可能である。

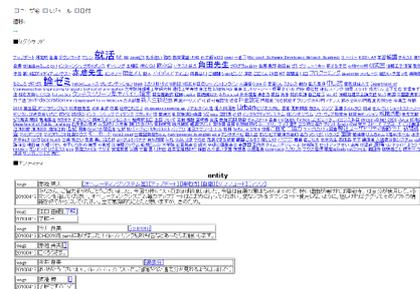


図 2: タグクラウド

4 利用例

例えば「ヒューマンインタフェース」に関連する情報を検索するときの本システムの利用方法を説明する。

検索画面で、「ヒューマンインタフェース」を指定すると、「ヒューマンインタフェース」のタグがついたエンティティの集合が取得される。同時に「ヒューマンインタフェース」との類似度から「インタフェース」や「interface」のサブ検索が行われ、これらのタグのついたエンティティの集合も取得される。検索結果は様々な表示方法を選択できる。

エンティティの集合に含まれるタグの出現頻度でタグクラウドとして表示や、出現頻度と類似度によるランク付けを行い関連しそうなエンティティを上位に表示、また他のエンティティとの関連をネットワーク構造などで表示することが可能である。

5 おわりに

本研究では、少グループ向けの文書情報共有システムを実装した。

今後は、類似度の計算方法を変えたり、メンバーの利用履歴などから、ユーザ個人に対するサジェストや検索をより正確にしていきたい。また、文書だけでなく写真やビデオなど様々なメディアに対応させ、より多様な検索ができるようにしたいと考えている。

参考文献

- [1] Michael Cashen, 角田博保, 赤池英夫. 少人数のためのコミュニケーション支援システム: Urban. インタラクション 2004. 2004.
- [2] Kerstin, B., Claudiy S. F., Wolfgang, N., Reluca, P.: Can all tags be used for search?, Proc. the 17th ACM conference on Information and knowledge management, pp.193-202,(2008).
- [3] Yahoo!JAPAN デベロッパーネットワーク, Yahoo! JAPAN, (<http://developer.yahoo.co.jp/>).