

構造化アノテーションを用いた知識再利用性の高い電子掲示板

松本 和之[†] 長尾 確^{††}

[†] 名古屋大学大学院 情報科学研究科 ^{††} 名古屋大学 エコトピア科学研究機構

1 はじめに

電子掲示板は知識の宝庫であるが、そのままでは再利用が困難である。従来型の掲示板ではプレーンテキストのみによる投稿が許され、単純なテキストマッチングによる検索程度しかできないため、掲示板から知識を効率よく抽出することが困難である。そこで本研究では、投稿時に投稿者が自分の投稿に対して構造化アノテーションを付与することによって、投稿データの再利用性を高める仕組みを提案する。

例えば、商品評価に特化した掲示板やアンケートに特化した掲示板等は存在するが、利用方法はそれぞれのシステムから提供されたものに限定される。本研究では、どのような用途にも利用でき、投稿データの利用法にまで自由度の高さを求めるシステムを目指した。

本研究は、機械学習を用いて電子掲示板から情報を抽出する手法 [1] とは異なり、始めから投稿を構造化しておくことによって投稿データを効率的に利用しようとする試みである。

2 投稿クラスを用いた構造化アノテーション

まず「投稿クラス」と呼ばれるデータを定義して、スレッドを作成するときや投稿を行うときにどのような構造で返信して欲しいのかを規定する。例えば「飲食店評価」クラスを持つプロパティには「店名」「ジャンル」「平均予算」「味の評価値」などがある。投稿時、XML 形式で記述された投稿クラスの定義からそれに対応した投稿フォームが自動生成されるので、投稿者はこのフォームに従って入力するだけで構造化された投稿が可能となる(図1)。このように、投稿内容に対してプロパティを対応させることにより構造化された投稿を行う本研究の仕組みを、構造化アノテーションと呼ぶ。

全ての投稿クラスは、デフォルトの投稿クラスから派生したクラスとして捉える。また他のユーザが作成したクラスを拡張し、性質を継承させたサブクラスをユーザが定義できるため、ユーザが収集したい情報に

図 1: 自動生成された飲食店評価の投稿フォーム

合わせて柔軟にプロパティを変えられる。ユーザ定義する際、プロパティ名、各プロパティが持つ値の型、各プロパティの投稿フォーム等を指定することができる。

3 投稿の絞り込み検索

蓄積された投稿は構造化アノテーションを用いて絞り込み検索を行うことができる。すなわち、図2のように各プロパティが持つ値を利用した、高度な絞り込み条件を指定できる。例えば「飲食店評価」クラスの投稿で「平均予算が1000円以下の中華の店」といったような条件で投稿を絞り込むことができる。また、プロパティの値が持つ型によって指定できる条件の種類が変化する。例えば、文字列型の値を持つプロパティに関する条件は、指定した文字列と「完全一致する」「含む」などの条件を指定できるが、数値型の値を持つプロパティでは、指定した値との大小比較などによって絞り込み条件を指定する。

絞り込みルーチンはユーザ定義することも可能である。ユーザ定義する際には、絞り込みルーチンを Perl スクリプトで記述する。このスクリプトはサブルーチンの形式で記述し、戻り値として投稿のリストを返すようにしておくことにより実現される。以上のようにして、ユーザが任意の絞り込み条件を設定することができ、また記述したスクリプトを他のユーザが利用したり拡張したりすることも可能である。

さらに、投稿クラスの継承によって、複数クラスにまたがった検索もできる。例えば、「営業時間」という

An Online Communication System with High Knowledge Reusability Using Structured Annotation

[†] MATSUMOTO, Kazuyuki(matsumoto@nagao.nuie.nagoya-u.ac.jp)

^{††} NAGAO, Katashi(nagao@nuie.nagoya-u.ac.jp)
Graduate School of Information Science, Nagoya University
(†)
EcoTopia Science Institute, Nagoya University (††)
Furocho, Chikusa-ku, Nagoya 464-8603, Japan

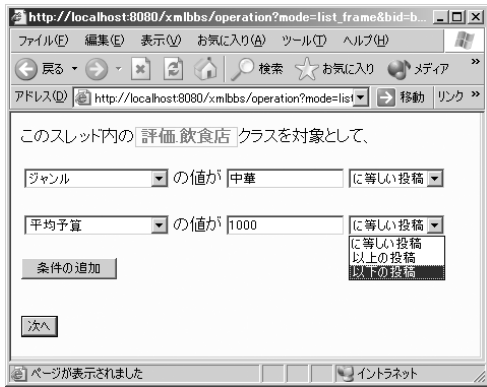


図 2: 絞り込み条件の指定

プロパティを持つ「店舗情報」クラスのサブクラスとして「飲食店情報」「駐車場情報」「アミューズメント施設情報」があるとす。全てのサブクラスは、その親クラスの属性「営業時間」が継承されているため、店舗情報クラスで「8:00以降まで営業している店舗」という条件で検索することによって、「飲食店情報」「駐車場情報」などといった異なるクラスであるにも関わらず、それらの中から条件に合った店舗情報を検索することができる。

4 オペレーションによる知識発見

蓄積された投稿を処理し何らかの結果を出す操作をオペレーションと呼び、単純なアンケートの集計やマーケティングなど、様々な知識発見のために利用できる。図3は、オペレーション実行画面の一例であり、ユーザ定義された「平日うまいもん巡り」というオペレーションが実行され、下部フレームにオペレーション結果が表示されている。なお、オペレーションは掲示板全体やスレッド内の投稿全体の他、先に述べた絞り込み検索の結果に対しても適用可能である(図4)。

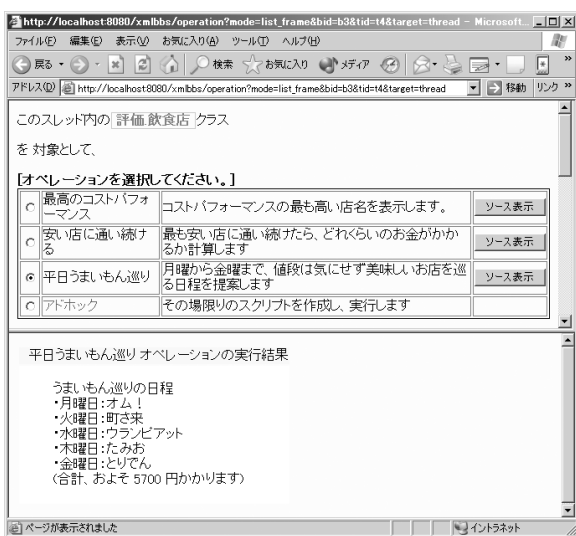


図 3: オペレーション実行画面の一例

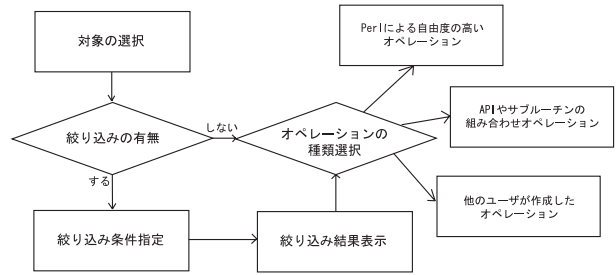


図 4: 絞り込み・オペレーションの流れ

ユーザによるオペレーションの定義は3通りの方法で行うことができる。もっとも自由度が高いのはPerlスクリプトを用いて記述する方法であり、Perlに関する知識があれば、投稿内容・日付・投稿者などの情報を自由に用いてユーザの求める処理を書くことができる。もっとも自由度が低く簡単なものは、他のユーザが作成したオペレーションの中から選択し実行する方法である。これらの中の方法は、既存の標準APIやサブルーチンをGUI上で組み合わせて構成する方法である。本システムには標準APIが実装されており、数値プロパティの平均・分散・標準偏差などを求めるものや、文字列の形態素解析結果を返すものがある。またPerlスクリプトを用いて引数付きサブルーチンを記述・登録することもできる。これらのサブルーチンを組み合わせることにより、比較的容易にオペレーションを構成できる。このように3種類の方法を用意することにより、知識の豊富なユーザからそうでないユーザまで幅広くオペレーションを利用できる。

5 まとめと今後の課題

本システムによって、構造化アノテーションを用いることで、電子掲示板上で様々な絞り込みやオペレーションを実行でき、その結果から知識発見が可能となった。このように、どのような用途にも利用でき、投稿データの利用法にまで自由度の高さがあり、さらに幅広いユーザ層に対応するシステムを構築した。

今後、グラフ等を用いたよりグラフィカルなオペレーション結果表示インタフェースや、Perlスクリプトを用いて絞り込みやオペレーションを記述したときのエラー処理やセキュリティ問題、ネット上の複数の電子掲示板システム間の投稿データやオペレーションの共有などについて検討する必要がある。

参考文献

- [1] 藤村滋ほか, “機械学習による電子掲示板からの評判情報抽出,” 情報処理学会第65回全国大会講演論文集(4), pp.451-452, 2003.
- [2] 松村真宏ほか, “オンラインコミュニティにおけるチャンス発見,” 人工知能学会誌 Vol.18 No.3, pp.295-300, 2003.