

# ユビ de コミハサンダー

## Ubiquitous Metadata Scouter – Semantics-based Information Search

稲葉 真純<sup>\*1</sup> 長野 伸一<sup>\*1</sup> 川村 隆浩<sup>\*1</sup> 長谷川 哲夫<sup>\*2</sup> 大須賀 昭彦<sup>\*2</sup>  
 Masumi INABA Shinichi NAGANO Takahiro KAWAMURA Tetsuo HASEGAWA Akihiko OHSUGA

<sup>\*1</sup> (株)東芝 研究開発センター      <sup>\*2</sup> (株)東芝 ソフトウェア技術センター  
 Research & Development Center, Toshiba Corp.      Software Engineering Center, Toshiba Corp.

In this paper, we introduce a service that is a semantics use case in ubiquitous computing environment. Ubiquitous Metadata Scouter searches through blogs on the internet when the cellular phone camera scans the product bar code. By referring to ontologies, it summarizes blog contents. It shows reputation for product and information on other related products, and some selected blog contents. Semantics extracts only worthy information to support user decision making and behavior in ubiquitous computing environment.

### 1. はじめに

ユビキタスコンピューティング環境における情報提供サービスでは、ユーザの状況や嗜好に応じた情報を提示することが求められる。例えば映画館にいればオススメ映画の情報を、ドライブ中であれば周辺の交通情報を地図で案内することが考えられる。従来のデスクトップコンピューティング環境では、広い画面と高速な通信回線を利用して、多くの情報から必要な情報を探出すことができた。しかし、ユビキタス環境の代表端末である携帯電話は、画面が狭く通信回線も細い。また、外出時などのユーザの状況によっては、時間をかけて情報を吟味することが難しい。ネットの情報を利用して、現実世界でのユーザの意思決定や行動を助けるには、必要な情報をピンポイントで提案する必要があるだろう。そこで、メタデータとオントロジーを情報検索に活用し、意味情報(セマンティクス)に基づいて抽出した有用な情報のみをユーザに提示する。本論では、ユビキタス環境におけるオントロジー活用の事例としてユビ de コミハサンダーを紹介する。

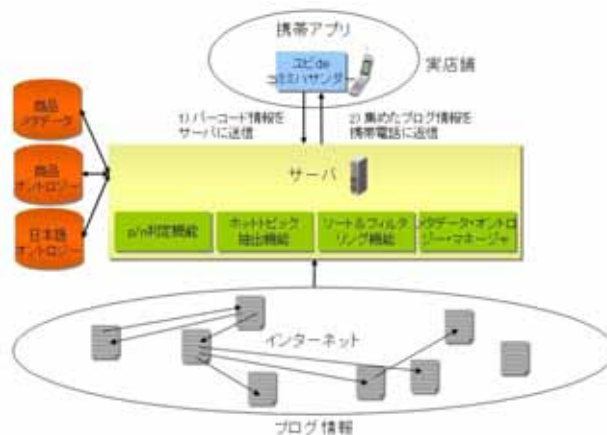


図1 ユビ de コミハサンダーの構成

### 2. ユビ de コミハサンダー

近年、商品に関する感想や使用経験をブログで公開するユーザが増えている。また、商品購入の際には、ブログの意見を比較検討の参考にすることもある。ユビ de コミハサンダーは、携帯電話のカメラで商品のバーコードを読み取ると、ネットワーク上のメタデータやオントロジーを活用して、その商品のブログでの評判や話題になっている関連商品などの口コミ情報をリアルタイムでユーザに提示するシステムである。ユビ de コミハサンダーのシステム構成を図1に示す。例えば、ユーザがある本のバーコードをスキャンすると、バーコードから、本のタイトルや著者、出版社などの情報を取得し、本の感想を述べたブログを収集する。収集したブログの内容を解析し、評判情報を集計するポジティブ・ネガティブ判定機能や、話題の関連商品を提示するホットピック抽出機能を持つ。また、ソート&フィルタリング機能によって、有用なブログのみを表示する。これらの機能により、ユビキタス環境において誰でも簡単にネットの情報を利用することができる。ユビ de コミハサンダーの出力例を図2に示す。



図2 ユビ de コミハサンダーの出力例

連絡先: 稲葉 真純, masumi.inaba@toshiba.co.jp  
 (株)東芝 研究開発センター 知識メディアラボラトリー

## 2.1 ポジティブ・ネガティブ判定機能

ポジティブ・ネガティブ(p/n)判定機能は、テキスト要約技術のひとつとして認知されている[立石]。本システムでは、収集したブログの文章に対して形態素解析、構文解析を行うと共に、トラックバックやコメントなどのブログ特有の構造関係を利用する。ブログのメタデータである RSS からブログコンテンツ間の相関関係を取得し、各コンテンツに対して重み付けを行う。例えば、トラックバックによって形成されるブログコミュニティの中で、オピニオンリーダーの役割を果たすブログは重視するなどの手法がある。同時に、現在約 10 万件の概念を有する日本語オントロジーを判定に利用する。日本語オントロジーの is-a 階層を参照して、表現の強弱などを考慮する。また、part-of 関係を参照することで、直接的にポジティブまたはネガティブではない表現についても判定を行う。さらに、instance-of の関係から、同一概念を表現する言葉の多様性を吸収する。これらの重み付け情報を用いて、ポジティブ・ネガティブのポイントを算出する。

## 2.2 ホットピック抽出機能

ホットピック抽出機能は、キーワードの出現頻度を利用したマイニング技術として研究されている[藤木]。本システムでは、ポジティブ・ネガティブ判定機能と同様にブログコンテンツ間の相関関係を利用するほか、商品の概念関係を記述した商品オントロジーを参照する。商品オントロジーでは、概念間の距離やつながりの疎密によって、ユーザが関心を持った商品に関連の深い商品を複数選択する。選ばれた商品と、ブログから得られるキーワードのマッチング処理を行うことで、より関連性の高い話題の商品を提示する。データベースであるオントロジーと、日々新たに生み出される動的なブログを組み合わせることで、バーコードをスキャンした時点での評価を得ることができる。現在は約 140 万件の商品メタデータと約 40 万件の商品オントロジーを有するが、新商品や商品改定などに対応し、随時データの追加や見直しを行っている。

## 2.3 ソート&フィルタリング機能

ソート&フィルタリング機能は、有用と思われるブログのみを抽出して表示する。ソート機能ではブログの相関関係を用いて、多くのトラックバックを受けている意見や、ポジティブ・ネガティブが明確に表明されている意見を重視する。商品の特徴や使用感が伝わる文章は読み応えがあり、予備知識が乏しい商品分野では、商品を比較するための観点をすることもできる。また、近頃では宣伝を目的としたスパムブログが増加しており、不要なブログを取り除くフィルタリング機能が必須である。

## 2.4 メタデータ・オントロジー・マネージャ

本システムでは、RDF で記述した商品メタデータと、OWL 形式の商品オントロジー及び日本語オントロジーを利用している。商品メタデータは、バーコードや商品名、製造メーカーなどの商品に関する詳細な情報を含む。一方、商品オントロジーは商品ジャンルを表すクラスと、固有の商品を表すインスタンスで構成される。また、日本語オントロジーは、モノに対する評価表現をまとめたものである。これらのデータは、メタデータ・オントロジー・マネージャで管理する。マネージャは、データファイルを構文に沿って解析し、取得した情報をデータベースに格納する。そして、条件に合う情報を取り出す検索機能を持つ。一連の機能を

有するメタデータ・オントロジー API としては Jena [Jena] や OWL API [OWL API] が知られており、充実した操作環境と検索機能を提供している。しかし、これらの API では、本システムのような数百万件に及ぶ大量のデータを高速に処理することはできない。データの増加に従って、読み込み速度や検索性能の低下、蓄積メモリ量の増大などの問題が顕著になるためである。本システムのメタデータ・オントロジー・マネージャは、これらの問題を解決し、ユーザが要求するリアルタイム検索に対応する。データはツールなどを用いて半自動的に生成することを想定しており、データの解析や管理方法を工夫することで読み込み速度の向上を図る。マネージャは事前に、メタデータ管理情報のペア(バーコード、メタデータ・ファイルパス)と、オントロジー管理情報のペア(商品名、オントロジー・ファイルパス#親クラス)をインデックスとしてメモリ上に蓄積する。メタデータ検索が要求されると、インデックスからメタデータ・ファイルパスを取得し、適時解析して必要な情報を取り出す。メタデータはバーコードによって検索され、本システムでは商品名を探し出す。続いて、商品名でオントロジーを検索し、オントロジー・ファイルパスと親クラスを取得する。同一ファイル内で同一の親クラスを持つインスタンスを収集し、スキャンした商品に関連の深い商品群として列挙する。各データには ID としてバーコードを付与し、データの重複を防止するほか、逐次実施されるデータの追加や削除、既存データに対する修正などのメンテナンスに配慮する。また、日本語で記述されたメタデータやオントロジーを円滑に処理するため、使用禁止文字の変換や文字コード処理を行う。

## 3. まとめ

本論では、実店舗でブログ情報を利用するためのオントロジー活用事例を紹介した。

本システムは 2006 年 3 月より家電量販店、大型書店にて実証実験を行っている。実験は一般ユーザに携帯電話を渡し、関心のある任意の商品をスキャンしてもらう方法を使った。今後は実験の結果を元に、機能の洗練化や新たな機能の開発を行う予定である。レスポンス速度の向上や、情報精度の向上に向けた改良を重ねていく。また、オントロジーの容易なメンテナンスを実現することも課題である。

広大なネットワーク上のリソースから情報を収集し、ユーザにとって未知の情報や、現実世界での行動を支援するサービスを発見することが求められている。収集した膨大な情報に対して様々な観点から分析を行うことで、多様化するユーザの嗜好やニーズに対応できると考える。ネットワークの情報を容易に利用できるシステムによって、現実世界とネットワークの連携を図っていきたい。

## 参考文献

- [立石] 立石健二, 福島俊一, 小林のぞみ, 高橋哲朗, 藤田篤, 乾健太郎, 松本裕治, Web 文書集合からの意見情報抽出と着眼点に基づく要約生成, 情報処理学会研究報告, 自然言語処理研究会, 2004-NL-163, pp.1-8, September 2004.
- [藤木] 藤木 稔明, 南野 朋之, 鈴木 泰裕, 奥村 学, document stream における burst の発見, 情報処理学会研究報告, 2004-NL-160, pp.85-92.
- [Jena] Jena: <http://jena.sourceforge.net/index.html>.
- [OWL API] Sean Bechhofer, Phillip Lord, Raphael Volz, Cooking the Semantic Web with the OWL API, 2nd International Semantic Web Conference, 2003.