

# ネットとリアルを結び付けるオントロジー技術 “ユビ de コミミハサンダー”<sup>TM</sup>

Ubiquitous Metadata Scouter — Ontology Connecting Blogs to Daily Life

川村 隆浩

長野 伸一

稲葉 真純

■ KAWAMURA Takahiro

■ NAGANO Shinichi

■ INABA Masumi

“ユビ de コミミハサンダー<sup>TM</sup>”とは、携帯電話からインターネットを検索するシステムである。携帯電話のカメラで商品のバーコードをスキャンすると、インターネットから商品のメタデータ<sup>(注1)</sup>を取得し、関連するブログを収集して、その商品の口コミ情報を抽出する。特に、オントロジー<sup>(注2)</sup>を参照しながらブログの中身を解析し、ユーザーの感性に沿った評判情報を出力することに特徴がある。

2006年3月には家電量販店と大型書店で実証実験を行い、精度面と速度面の検証を行った。今後ますます増えていく、ユビキタス環境でのインターネット利用に対する一つのアプローチである。

Ubiquitous Metadata Scouter is a mobile service for searching the Web from cellular phones. When a user takes a photo of a product's bar code using a cellular phone camera, the service first retrieves the corresponding metadata for the product (data identifying the product such as the name and manufacturer) from the Internet, then collects the related blogs. The main feature of the service is that the contents of each blog are analyzed using ontology, to indicate the overall reputation of the product. Ontology consists of the specification of a conceptualization in order to classify concepts relating to a certain domain.

Toshiba conducted verification tests of the service at a consumer electronics store and a bookstore in Tokyo in March 2006. This service holds promise as an Internet search method in the expanding ubiquitous computing environment.

## 1 まえがき

ユビキタスコンピューティング環境においては、誰でも簡単にネットの情報を現実の世界と結び付けて利用できることが求められる。リアル連携とも呼ばれるこうしたニーズは、例えばユーザーが映画館にいれば映画の情報を、ドライブ中であれば周辺の名所を提示するなどが典型的な例である。従来のデスクトップコンピューティング環境では、広い画面と使いやすいインタフェース、そして高速な通信回線があるため、情報検索システムはユーザーに対して一度にたくさんの情報を出すことができ、ユーザーも何度も試すことができた。今や、検索エンジンが出力する100件近いリストの中を何度もクリックして、所望の情報を探し出すことに抵抗を感じる人も少ないだろう。

しかし、ユビキタス(現時点では、まだモバイル)環境の代表端末である携帯電話では、画面は狭く、操作は限られ、回線もまだまだ細い。そのため、もっと簡単にネットの情報を利用するためには、操作をより自動化し、ユーザーが必要な情報だけを提示することが必要である。

(注1) データの内容を説明する付加情報。例えば、商品メタデータの場合は商品名や製造元など。

(注2) 特定分野の概念をまとめた語体系。

そこで、われわれは近年利用が盛んなメタデータとオントロジーを活用し、意味情報(セマンティクス)に基づいてユーザーに有用と思われる情報やサービスを抽出して提示する技術の研究・開発を進めている。ここでは、ユビキタス環境における情報検索技術に活用した事例として、“ユビ de コミミハサンダー<sup>TM</sup>”について述べる。

## 2 ユビ de コミミハサンダー<sup>TM</sup>とは

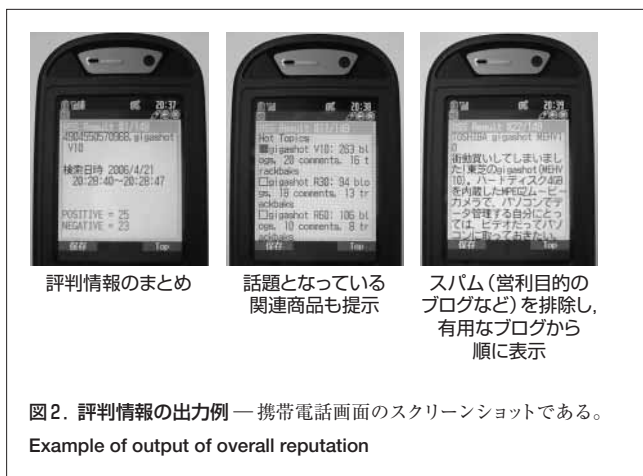
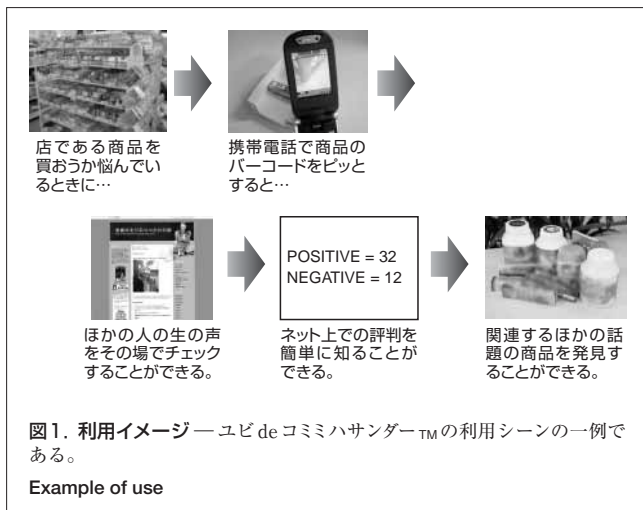
近年、商品に関するニュースリリースをいち早く取り上げたり、購買時の比較検討や使用経験などをユーザー独自の視点でブログに公開するユーザーが急増している<sup>(1)</sup>。一方、ユーザーは商品購入の際に、他人のブログに書かれた商品に関する情報を収集し、比較検討を行って購入を決定していることが少なくない。更に、その比較検討した結果をユーザーがブログで公開することもあり、口コミ情報がますます増えてきている。

ユビ de コミミハサンダー<sup>TM</sup>とは、携帯電話のカメラで商品のバーコードをスキャンすると、インターネットから商品のメタデータを取得し、関連するブログを収集し、その商品の口コミ情報をリアルタイムでユーザーに提示するものである。例えば、ユーザーがある本のバーコードをスキャンすると、システムはそのバーコードから、本のタイトルや出版社、著者

といった情報を取得し、書評などを述べているブログを収集する。同時に、ブログのメタデータを参照しながらブログの中身をオントロジーを用いて解析し、商品の評判情報を提示する(ポジティブ・ネガティブ判定機能)。また、ほかの関連商品で話題となっているものを提示する(ホットピック抽出機能)。そして、最後に有用と思われるブログをいくつか表示する(ソート&フィルタリング機能)。ユビ de コミミハサンダー™の利用イメージを図1に、また、出力例を図2に示す。

このシステムにおける東芝の独自技術は以下の2点である。

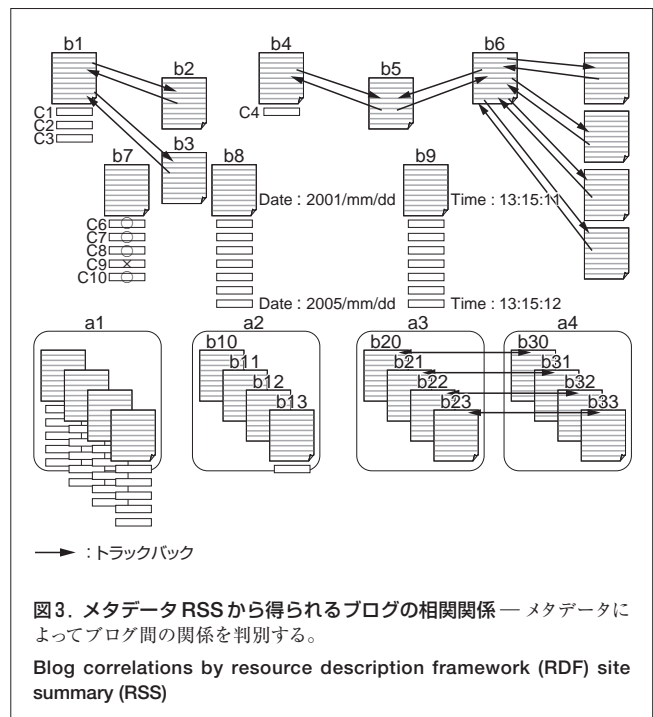
一つ目は、ブログのメタデータ RSS1.0 (RDF (Resource Description Framework) Site Summary)<sup>(2)</sup> から得られるブログエントリ(ブログ内の個々の記事)間の相関関係を利用し、各エントリの重要さを重み付けし、評価をバイアスしている。従来の研究では、互いに独立した文書の集まりを前提としている場合が多く、ユーザーがウェブ上のリンクをたどりながら様々な意見に出くわす状況を想定していない。



しかし、実際には、例えばブログにおける意見において、多くのトラックバック<sup>(注3)</sup>によって賛意が付けられているものと、一つもトラックバックが付いていないものとはユーザーの心理に与える影響は異なる。

そこで、このシステムでは対象文をブログに限定することで、そこに付与されたRSSからブログ間の相関関係を取得し、活用している。具体的には、以下のような戦略を用いている。ブログの相関関係の例を図示したものを図3に示す。

- (1) 単なるコメントよりも、トラックバックを重視する(出自を明らかにしている: b2やb3)
- (2) 意見をサポートする著者の異なる意見(トラックバック又はコメント)の量に応じて重みを付ける(多くの人に支持されている: b6)
- (3) ほかの関連商品についても意見を述べている著者の意見は重視する(エキスパート、同一観点からの比較が期待できる)
- (4) トラックバックやコメントの流れにおいて、反対の中の反対よりも賛成の中の反対を重視し、同様に、賛成の中の賛成よりも反対の中の賛成を重視する(勇気ある発言: b7のc9)
- (5) 元のコンテンツが古いにもかかわらず、長く意見が付いている意見は重視する(パイオニア: b8)
- (6) 初めの意見から最新の意見までの数を時間で割った値が大きい場合はそれぞれの意見に重みを付ける(加速度が大きい: b9)
- (7) 意見数が平均的に多い著者の意見は重視する(オピニオンリーダー: a1)



(注3) ブログエントリに対してリンクを張ったことをリンク参照先へ通知する機能で、プログラマーのコミュニティ形成の基本。

- (8) 意見の付いていないコンテンツは重みを下げる(広告などを排除: b10-12)
- (9) 意見の付いていないコンテンツの多い著者の意見は下げる(広告業者などを排除: a2)
- (10) 元コンテンツの著者と意見を付けた者の二人が相互に意見交換をしている場合は重みを増す(議論が深まっている: a3とa4)
- (11) 最後に、ある調査<sup>3)</sup>によれば、ブログでは70%が良いことを述べる傾向が見られる。そこで、反対にネガティブな意見があれば、著者の強い意志を考慮して評価をバイアスする。

むろん、それぞれの戦略に基づく重み量のチューニングが、抽出精度を大きく左右する。そのため、次章における実証実験を通してチューニングを行っている。

独自技術の二つ目は、日本語オントロジーの概念体系を参照して表現に強弱を適用し、単に商品に対する良い悪いといった表現をカウントするのではなく、表現の程度に応じて点数を付けている。また、構文解析を行うことで、同じ表現でも意味が反転する場合を考慮している。例えば、車の“パ

フォーマンスが高い”は良い意味となるが、“値段が高い”は悪い意味と判別される。また、同一概念の異なる表現も吸収される。オントロジーの例を図4に示す。

このシステムでは、特定の商品に関するブログを最大100件収集し、評判抽出を行っている。これにより、単に文書集合を正確に判定した結果ではなく、ユーザーがブログをブラウズしながら得る印象値により近い評判情報を得ることができる。次章において、この点について検証する。

### 3 実証実験と評価

このシステムは評価版の開発を完了し、2006年3月に家電量販店、大型書店の協力を得て実証実験を行った。実験では、まず被験者19名にユビ de コミミハンダー™をインストールした携帯電話を渡し、実際に店舗で気になる商品を手にとってもらい、バーコードをスキャンして評判情報を読んでもらった。そして、実験会場に戻った後、こちらで用意したアンケートに沿っていくつかの質問に答えてもらった。なお、被験者候補にはあらかじめプロフィールを提出してもらい、年齢や性別、インターネットに関する知識などの観点から数名ずついくつかのグループに分けて採用した。今回のシステム構成を図5に示す。

以下に、ポジティブ・ネガティブ判定(以下、p/n判定と略記)の精度面の評価の一部を示す。表1のユーザーの印象値は、今回の被験者らが対象商品に関する複数のブログスレッドを実際に読んだ結果、商品に関してポジティブ又はネガティブとして得た印象値の平均である。

一方で、表1の専門家によるp/n分析は、言語解析などを研究する人間が同じブログスレッド内の個々のエントリを

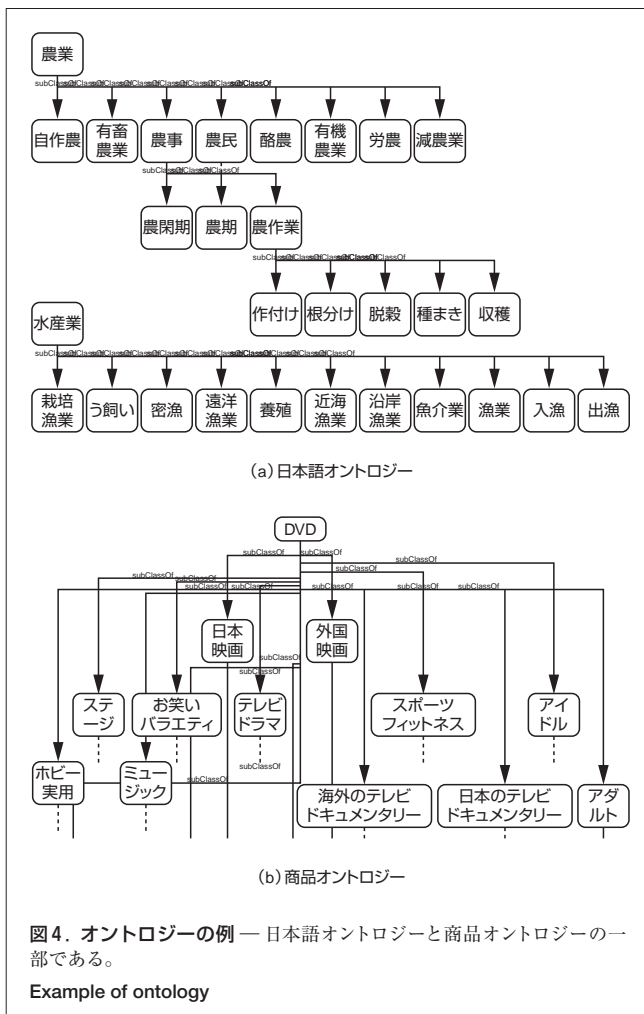


図4. オントロジーの例 — 日本語オントロジーと商品オントロジーの一部である。

Example of ontology

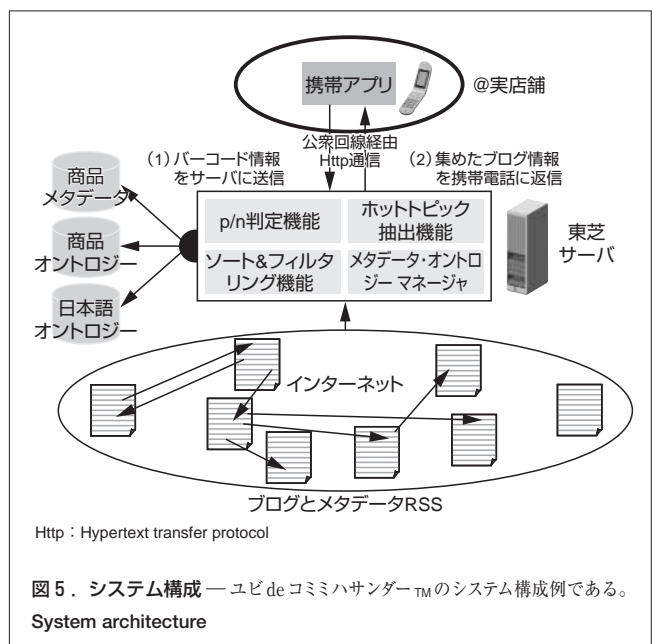


図5. システム構成 — ユビ de コミミハンダー™のシステム構成例である。

System architecture



表1. p/n 判定の実験結果

Results of positive/negative (p/n) determination

対象商品	ユーザーの印象値	
	ポジティブ (%)	ネガティブ (%)
商品1 (DVD)	86.8	13.2
商品2 (書籍)	95.0	5.0
商品3 (HDDレコーダ)	81.8	18.2
専門家によるp/n分析		
	ポジティブ (%)	ネガティブ (%)
商品1 (DVD)	62.5	37.5
商品2 (書籍)	71.4	28.6
商品3 (HDDレコーダ)	28.6	71.4
p/n 判定の出力		
	ポジティブ (%)	ネガティブ (%)
商品1 (DVD)	86.4	13.6
商品2 (書籍)	91.7	8.3
商品3 (HDDレコーダ)	100.0	0.0

p/n : positive and negative

じっくりと読んでいき、賛成者と反対者(賛成でないものを含む)をカウントした割合である。これらを見比べることによって、実際にブログを読んで得られるユーザーの感性と、実際の賛否の数とは必ずしも一致しないことが確かめられた。これは2章で述べたわれわれの仮説を裏づけており、ユーザーにとってのp/n判定は単純な評価表現の総数ではなく、表現の程度やブログエントリ間の関係に影響を受けているだろうことが推察される。

次に、今回のp/n判定処理について説明する。まず、ブログ検索ロボットにより商品名(又はその一部)が含まれるブログエントリを最大100件収集する。そして、形態素解析と構文解析を行い、商品名と係り受け関係のある評価表現を抽出する。そのうえで、2章で述べた二つの独自技術を適用している。ここでは、まず独自技術を適用せずにp/n判定を行った。そして、その結果を表1の専門家によるp/n分析結果と比較し、ベースとなるp/n判定処理自体の正解率を求めたものを表2に示す。

ただし、1エントリ内でほめていたりけなしていたりする場合は、各表現に+点と-点をつけて合算している。HDDレコーダ以外は80%を超える正解値を示し、一定の精度が得られていることがわかる。HDDレコーダの正解率が悪い理由としては、業者による宣伝ブログが多く含まれていたため

表2. p/n 判定(独自技術なし)の専門家分析に対する正解率  
Accuracy of p/n determination compared to expert's analysis

対象商品	正解率 (%)
商品1 (DVD)	87.5
商品2 (書籍)	85.7
商品3 (HDDレコーダ)	57.1

である。表1の専門家はそれらを営利目的としてカウントしなかったが、p/n判定処理は宣伝かどうかを判断できずポジティブにカウントしているものが多かった。この点に関しては、今後宣伝ブログの排除機能の強化が急務であると言える。

更に、このp/n判定処理に2章で述べた二つの独自技術を加えてp/n判定を行った結果を、表1のp/n判定の出力に示す。この結果から、このシステムはユーザーの感性に近いp/n判定結果を得られていることがわかる。これにより、提案した改良方法に一定の有効性があることが確かめられた。

また、速度面では、現在、結果を表示するまでに約10~30秒ほどかかっている。ただし、実験用サーバはたかだかPentium 4, 3.2 GHz, メモリ1Gバイトのデスクトップパソコンであり、商品オントロジーは40万件、商品メタデータ140万アイテムを用いた。アンケートによれば、20秒を超えるとユーザーは遅く感じる事が判明したため、今後高速化を図っていく。

## 4 あとがき

オントロジーと呼ばれる、意味情報を用いてブログをリアル店舗で活用できるように要約する携帯電話向けサービスについて述べた。今後は、実証実験の更なる評価を基に、精度面からは適合率の向上、速度面からは特に複数同時アクセス時のレスポンス向上などの改良を図っていく。

## 文 献

- 総務省ブログ. SNSの現状分析及び将来予測.  
< [http://www.soumu.go.jp/s-news/2005/050517\\_3.html](http://www.soumu.go.jp/s-news/2005/050517_3.html) >,  
(参照 2006-06-16).
- D. Brickley, et al. RDF Site Summary (RSS) 1.0.  
< <http://purl.org/rss/1.0/spec> > , (accessed 2006-06-16).
- ITmedia Inc. “プログラマーの製品チェック、切り口は4類型”. ITmedia News.  
< <http://www.itmedia.co.jp/news/articles/0511/28/news070.html> > ,  
(参照 2006-06-16).



川村 隆浩 KAWAMURA Takahiro, Ph. D.

研究開発センター 知識メディアラボラトリー研究主務, 工博。主にセマンティック Web, ソフトウェアエージェントの研究・開発に従事。人工知能学会, 情報処理学会会員。Knowledge Media Lab.



長野 伸一 NAGANO Shinichi, Ph. D.

研究開発センター 知識メディアラボラトリー研究主務, 工博。主にセマンティック Web, ソフトウェアエージェントの研究・開発に従事。IEEE, 情報処理学会, 電子情報通信学会会員。Knowledge Media Lab.



稲葉 真純 INABA Masumi

研究開発センター 知識メディアラボラトリー。主にセマンティック Webの研究・開発に従事。Knowledge Media Lab.