

F-007

ブログを用いた Web サービス型情報共有・推薦システム「HARMO」の開発

HARMO : Information Sharing and Recommender System with Weblog based on Web Service

山下 晃弘[†] 川村 秀憲[†] 山本 雅人[†] 大内 東[†]
 Akihiro Yamashita Hidenori Kawamura Masahito Yamamoto Azuma Ohuchi

1 はじめに

Web上の情報資源は現在も急速に増加している。その多くは、映画、風景、ニュース、食品、電気製品、店舗など世の中のあらゆる「モノ」に関連し、情報の存在場所や表現形式も多様である。この膨大な情報の価値を高めてきたのが検索技術であり、ユーザが必要とする情報を効率良く提供できる環境を作り上げてきたといえる。

将来的には、各々の情報がオントロジを持ちユーザのニーズをシステムが意味レベルで理解して情報を提供するセマンティック Web が実現するといわれており、多くの研究が行われている [1]。ただし、その実現には技術体系の確立や標準化など多くの課題解決と環境の整備が不可欠で、インターネット全体の改革であるこの技術の普及には時間をかけて取り組む必要がある。

一方現実問題として、Web 情報の活用における利便性向上のニーズは高まるばかりである。キーワード検索の精度向上により格段に便利になったとはいえ十分とは言えず、検索にある程度の経験と知識が要求される場合や、興味や好みなどそもそもキーワードでの表現が困難な検索も考えられる。そのような課題から近年注目されているのがパーソナライゼーションである。これは、システムが様々な情報の中から個人適応的に選択して提供する技術である。

このような技術やサービスが今後も一層重要になる中、Web上の雑多な情報から、必要な情報をいかに利用しやすい形で整理しシステム内に取り入れるかが一つの課題である。そこで本研究では、世の中のあらゆるモノの情報をユーザ間で共有し、個人の興味や嗜好に添って提供する Web サービス型システム「HARMO (Hyperinformation Assembling Recommendation Management Organizer)」を開発した。

本システムでは世の中のあらゆるモノに関する情報を表現するために「HARMO カード」という情報単位を定義する。そして様々なシステム上で HARMO カードを利用するための Web サービスを構築した。本稿ではシステムの概要について述べた後、ユーザプロファイリング技術を用いた情報選択機能の実現や、ブログに適用することによる情報収集と共有について述べる。また、実

際にシステムを実装し試験運用をした結果について報告する。

2 システム概要

2.1 システム構成

HARMO は、全ての情報を管理し Web サービスとして提供するデータサーバと、その Web サービスを利用し様々なコンテンツをユーザに提供するインタフェースサーバに分けることが出来る [図 1]。

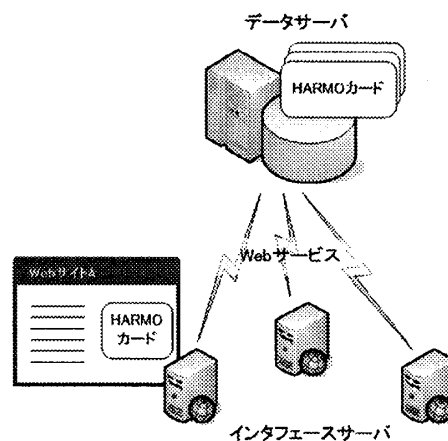


図 1: システム構成図

データサーバ上では、世の中のモノに関する様々な情報が一元管理され、外部からのサービス要求に対して情報を提供する。一方インタフェースサーバは具体的な Web サイトを構築するサーバであり、データサーバが提供する情報を活用してユーザ間コミュニティを形成したり情報提供サイトを構築する。インタフェースサーバは任意のドメインに設置が可能であり、後から自由に追加することができる。

2.2 HARMO カードの導入

本システムでは、全ての情報が HARMO カードという単位で扱われる。1 枚のカードは 1 つのモノを表現し、それに関する様々な情報はカード上で「基本情報」と「個人評価」に分類される。

[†]北海道大学 大学院情報科学研究科

基本情報は、モノに付随しユーザには依存しない情報であり、モノの住所や価格がその典型例である。情報の正確さが必要であり、全てのユーザ間で共有することが可能な情報である。情報提供者が多くの情報を収集するためには手間とコストが必要だが、wikipedia¹のようにユーザが持つ知識を集合させ、一つの巨大なデータベースを構築した事例もある。本研究ではこのような集合知を利用した情報収集を試み、多くのユーザが HARMO カードを参照し、情報を追加、修正することでシステム全体の充実を図る。

一方個人評価は、あるモノに対する個人に依存した情報で、主観的な評価や感想などである。この情報は個人の興味や嗜好を表現した情報であると捉えられ、パーソナライゼーションを実現する上で必要不可欠である。

インターネット上では個人の評価や感想を収集しているサイトも多く見かけるが、共有化されるべき情報と、個人の主観的・感覚的情報が混在してしまうと、誹謗中傷問題の一つの原因になりかねない。本システムでは、これらを明確に区別することで、必要な情報は共有化し、個人評価はパーソナライゼーションに活用する。

HARMO カードは、予めいくつかのカテゴリに分類している。これにはいくつかの理由がある。まず、HARMO カード上の基本情報を単にテキスト情報ではなくデータベースとして構築するために、カテゴリごとに項目を設定している。例えば、カテゴリの一つである飲食店の場合は、住所や電話番号が項目である。これによりカテゴリ内での情報の粒度を統一し、様々な Web サイト上で HARMO カードを参照した際に同一のフォーマットで表示するようにした。

また、カテゴリ分けをすることで、個人の興味や嗜好を元に情報を選択する際の比較対象を明確化できる。全くジャンルの異なるコンテンツ同士の比較は難しくユーザの興味に則した情報提供の正確さを低下させる恐れがある。本研究ではコンテンツのカテゴリを一つの属性と捉え、情報提供時にはあるカテゴリ内でのランキングを表示することにする。

加えてカード検索時の利便性向上が挙げられる。カテゴリを明確にしておくことでキーワード検索に加えカテゴリを指定した検索機能を実現した [図 2]。

2.3 XOOPS モジュール化

本研究では、インタフェースサーバの一つの構築例として XOOPS のモジュールを実装した。

¹<http://wikipedia.org/>

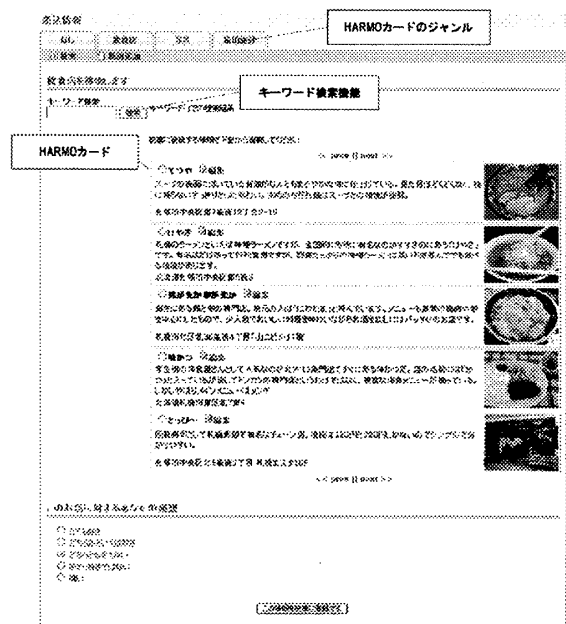


図 2: HARMO カードの検索 (飲食店カテゴリ)

XOOPS は近年ウェブサイト構築するフレームワークとして普及している CMS (Contents Management System) の一つであり、様々な機能をモジュール単位で追加拡張可能であることが特徴である。利点として、XOOPS 本体のインストールやモジュールの追加が容易であることや、ユーザ管理機能など本システムに必要な機能が予め準備されていることなどが挙げられる。既に XOOPS 上でウェブサイトを構築している場合は、本研究で構築したモジュールをインストールするだけでデータサーバ上に存在する共有コンテンツを利用した情報提供サイトやブログを構築できる。

2.4 ユーザプロファイリングと個人適応化

各ユーザが必要とする情報をシステムが適切に選択するためには、個人の興味や嗜好の情報をいかにして取得するかが問題である。この技術をユーザプロファイリング技術と呼び、近年多方面から様々な研究がなされている [2][3]。

ユーザプロファイリング技術は、ユーザの手間という観点から、直接興味や嗜好に関する情報を入力してもらう明示的手法と、普段の Web 閲覧時の挙動からシステムが自動的に情報を収集する暗黙的手法の 2 つに大別される。明示的手法の代表例としては、コンテンツに対する評価付けが挙げられ、現時点でも多くのサイトで取り入れられている。また、暗黙的手法の代表例としては、テキスト解析やアクセスログ解析、マウス操作などを利

用する方法であり、徐々に普及し始めている。一般的に、前者はより信頼性の高い情報が収集できるがユーザの手間が増えてしまい、後者はその逆である。どちらにしろ、できるだけ入力の手間を軽減しユーザが積極的に安心して興味や嗜好の情報を提供できる環境が必要である。

次に、収集した情報を元に情報を選択するアルゴリズムが必要である。そのための技術として推薦システム(Recommender system)が挙げられる。初期の推薦システムは、主にネットニュースなどを対象としたシステムとして開発された[4]。現在ではインターネット上のあらゆるコンテンツを始め、Amazon.comに代表されるようなオンラインショッピングサイトでも活用されており、推薦アルゴリズムに関しても様々な研究がなされている[5]。

推薦システムは、そのランキング方式から大きく2つに分類される。一つは扱う情報の内容に基づきランク付けを行うコンテンツベースフィルタリング(Content-based filtering)で、もう一つがコミュニティ内で同じ嗜好を持つユーザが好む情報を優先的に選択する協調フィルタリング(Collaborative filtering)である。HARMOカード上において、ユーザが入力する個人評価は、HARMOカードに対応した特定のモノに対する評価である。従って、ユーザが持つ一般的な興味や嗜好の度合いを利用するコンテンツベースフィルタリングよりも、ユーザ間の嗜好の近さから提供するコンテンツを決定する協調フィルタリングのほうが適しており、情報選択アルゴリズムとして採用する。特に、本稿ではユーザ間の嗜好の近さを相関係数を用いて表現する「相関係数法」を採用することとした。

本システムにおけるランキング機能は全てデータサーバ上に実装し、Webサービスを介して利用される。これにより、様々なWebサイト上でパーソナライゼーションが可能になると共に、将来的な機能拡張が容易であることやランキングアルゴリズムの変更や改善の際も容易に対応できる。

2.5 ブログの利用

近年ブログが急速に普及している。ブログ上には、日常の出来事から専門的な話題まで多種多様であり、ユーザの主観的な観点に基づいて自由に記述されている。ブログ記事の中には、コンテンツに関する様々な情報が記載されているものも多く、その増加傾向は今後も衰える気配が無い。ブログ検索サイトの米Technoratiの調査²によれば、2006年第4四半期に世界中のブログで最も多

く使われた言語は日本語であり、このことから日本での著しい増加傾向が伺える。

また、ブログを対象としたサービスも徐々に多様化している[6]。様々なコンテンツや機能をパーツとしてブログに組み込み、ユーザが外部の情報をブログで利用可能になりつつある。このような状況から、ブログはますます個人に特化し、個人の趣味や嗜好を反映するツールになると考えられる。

しかし現在普及しているブログシステムにおいてユーザが発信する情報はブログサイトとして一般に公開されるのみである。ブログの記事内に存在する情報は、ユーザプロフィールに利用可能な個人評価や他のブログユーザと共有可能な基本情報が多く含まれており、これらを活用できれば新たな情報流通の仕組みを実現できる可能性がある。

そこで本研究では、あらゆるモノに関する基本情報や個人評価を収集するためにブログの利用を試み、システムを構築した。具体的には、記事内にHARMOカードを挿入することで、他のユーザとの情報共有が可能なブログシステムを構築し、その上でブログの利用者からはHARMOカードを介して個人評価を収集する。またHARMOカード上の基本情報についても追加や修正を可能とし、ブログ利用者が日々記事を投稿する中で、システム上に情報が蓄積されていく環境を構築した[図3]。

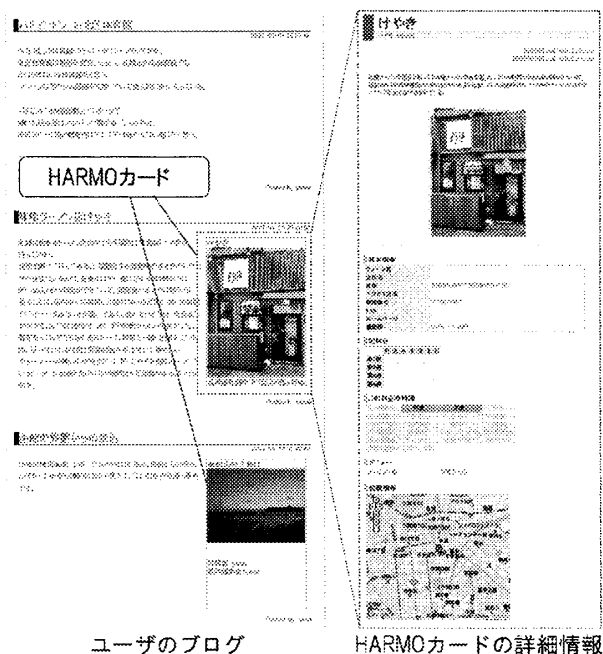


図3: HARMOカードを利用したブログシステム

本研究で構築したブログシステムのもう一つの特徴が、個人評価を利用したお勧めHARMOカード提供機能で

²<http://technorati.com/weblog/2007/04/328.html>

ある。ブログ記事において HARMO カードを利用する際に入力した個人評価は、投稿者に依存した主観的情報として扱われ、ユーザに適応した情報提供時に使用される。従って、普段ブログ記事を投稿する際に自分がどう感じたのかを明示的にシステムに伝えることで、システムがそのユーザの興味や嗜好を判断し、まだユーザの知らないコンテンツを自動ランキングして提供する。この機能により、日々の記事投稿から新たなコンテンツの発見へと繋げることができる。

3 実装と試験運用

図4は本システムのアプリケーション構成である。実装方針として、拡張性やメンテナンスの容易さを考慮し、オープンソース技術を積極的に取り入れることで低コストなシステムを実現した。Webサービスを利用する新たなプログラムを記述する場合は、PHPにより実装した HARMO ライブラリを使用するか、利用者自身が SOAP 通信を実装する必要がある。HARMO ライブラリを利用する場合は、専用のクラスを利用することで、HARMO カードを簡単に取得できるよう設計されており、誰でも拡張することが可能である。

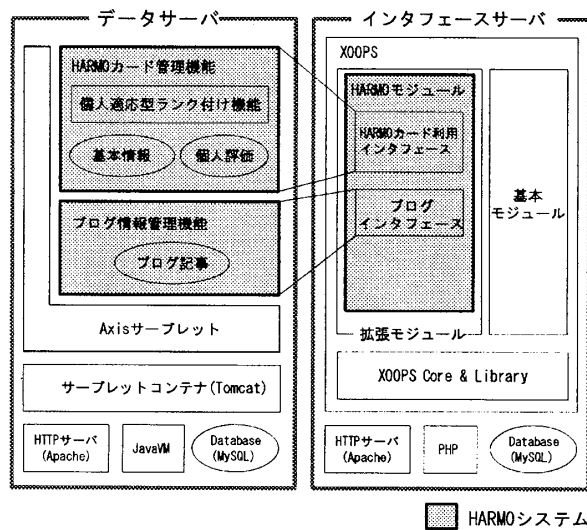


図4: アプリケーション構成図

構築したシステムの試験運用として本学学生 21 名にブログを開設してもらい情報の収集と共有化を試みた。その結果多くの記事に HARMO カードが挿入され、情報が共有化されると同時に基本情報の追加、修正や個人評価の収集を確認することができた。個人に適応したランキング機能については、まだデータが十分ではなく機能の評価には至っていないが、今後情報が充実するにつれ徐々に適切な情報提供がなされるものと思われる。

4 まとめ

Web 上では、あらゆる「モノ」に関する情報が増加し続けており、それらを適切に収集・整理し、個人の興味や嗜好に基づいて提供する技術の重要性が高まっている。本研究では、あらゆる「モノ」に関する情報を共有・推薦するシステム「HARMO」を開発した。その中で「HARMO カード」という情報単位を定義し、HARMO が管理する全ての情報は、ユーザ間で共有可能な基本情報と個人の興味や嗜好に依存した個人評価に分類した。共有化する情報は客観的に誰もが納得できる情報に限ると共に、個人評価はあくまで主観的な意見であることを明確にした上で情報提供やランキングに使用した。

HARMO は Web サービスとして様々な Web サイトに HARMO カードを介して情報を提供する。本稿では、その一例として XOOPS モジュールを実装し、ブログサイトを立ち上げると共に試験運用を行った。

今後は共有化された基本情報の正確さや更新頻度を解析すると共に、個人情報を利用したランキングアルゴリズムの評価を行う。また、ブログと HARMO カードが作り出すリンクネットワークは広い意味でユーザ間のコミュニティと捉えられるため、ブログ記事同士の関連性や HARMO カード同士の類似性にも着目していきたいと考えている。

参考文献

- [1] 武田英明: "Web の進化とエージェント, セマンティック Web", 情報処理, Vol.48, No.3, pp.229-235 March 2007.
- [2] 土方 嘉徳: "情報推薦・情報フィルタリングのためのユーザプロファイリング技術" 人工知能学会論文誌, Vol.19, No.3, pp.365-372, 2004.
- [3] 堀 幸雄, 今井 慈郎, 中山 堯: "個別化する Web : 検索キーワードの個別化" 情報知識学会論文誌, Vol16, No.4, pp.33-40, 2006
- [4] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. "GroupLens: An Open Architecture for Collaborative Filtering of Netnews" In CSCW '94: Conference on Computer Supported Cooperative Work (Chapel Hill, 1994), ACM, pp. 175-186.
- [5] Gediminas Adomavicius and Alexander Tuzhilin "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions" IEEE Transactions on Knowledge and Data Engineering, Vol.17, No.6, JUNE 2005 pp. 734-749.
- [6] 武田英明, 大向一輝: "Weblog の現在と展望—セマンティック Web およびソーシャルネットワークの基盤として—", 情報処理, Vol.45, No.6, pp.586-593, June 2004.