

利用者特性を考慮したディレクトリサービスの提案

仲澤誠志[†] 杉山達彦[†] 木下哲男^{††} 白鳥則郎^{††}

[†]東北大学電気通信研究所/情報科学研究科

^{††}東北大学電気通信研究所

本研究では、WWW における情報提供方式の一つであるディレクトリサービスにおいて、利用者の興味や傾向を獲得し、情報に対する索引付けや分類にフィードバックさせる手法を提案する。本手法では、利用者の特性および情報の特性をキーワードの集合で表現し、利用者が情報を閲覧する際にキーワードを相互に反映させることにより、利用者の特性の獲得だけでなく、情報への索引や分類をより適切なものに更新していくことが可能となる。

本稿では、従来のディレクトリサービスの問題点を取り上げ、本手法による解決法を述べている。また、利用者特性の獲得機構と索引の更新機能を設計して、本手法の妥当性を検討している。

A Proposal of New Directory Service based on User Profiles

Satoshi NAKAZAWA[†], Tatsuhiko SUGIYAMA[†], Tetsuo KINOSHITA^{††}
and Norio SHIRATORI^{††}

[†] Research Institute of Electrical Communication, Tohoku University
/ Graduate School of Information Sciences

^{††} Research Institute of Electrical Communication, Tohoku University

We propose a new directory service system which gets user profiles and feedbacks it to indexing and categorizing information on the World-Wide-Web. By representing the user profiles and information on the WWW in form of set of keywords, this system can get user profiles automatically. Moreover, such profiles can be used to maintain the index and category in more suitable form. In this paper, firstly, we discuss about problems of existing directory services and propose the solution against the problem. Then, we evaluate the effectiveness of our directory service system.

1.はじめに

近年のインターネットの普及にともない、WWW を用いて発信される情報の量が飛躍的に増大している。仕事・研究から趣味・娯楽まで、利用者が

必要とする情報の多くが WWW を用いて獲得できるという時代になってきている。しかし一方で、その必要な情報を膨大な情報の中から検索して獲得することは非常に難しく、利用者に常に負担を強いるという問題が生じている。このため、利用者

が必要とする情報に容易にたどり着くことを支援するサービスの必要性が高まってきた。そこで誕生したのが検索エンジンやディレクトリサービス(以下「サービス」)である。特に Yahoo![1]や Vector[2]に代表されるディレクトリサービスは、サービスの管理者が情報に対して索引付けを行い、さらに適切な分類を行うことで、WWW 上に存在する膨大な情報をサービスの利用者を使いやすいように整理することを目的[3]に運営されている。実際、WWW の利用者の大部分がこのようなサービスを利用[4]しており、その必要性は非常に高いものといえる。

しかし、従来のディレクトリサービスは、情報への索引付けや分類がサービスの管理者の主観に基づいて行われていたために、必ずしもサービスの利用者にとって使いやすい形式で情報の整理が行われているとは言えなかった。また、サービスを構築して保守する際には、管理者に多大な負担がかかるという側面がある。このような利用者と管理者の両方の視点からの問題を解決して、より使いやすいサービスを提供できるシステムを提案することが本研究の目的である。

そこで本研究では、サービスの利用者の興味や傾向を利用者特性(以下「プロフィール」という形式で表現[5]することにより、利用者要求の獲得やウェブマーケティングへの応用を図る。また、情報閲覧時に用いる索引(以下「インデックス」)に対して、多数の利用者のプロフィールからフィードバックを与えることにより、より適切な索引付けや分類を管理者によらず半自動化で行える仕組みを提案する。これにより、利用者が利用しやすい、また管理者が管理しやすいディレクトリサービスの提供が可能となる。

以降、2 章では、既存のサービスの問題点と提案手法による解決法について述べる。3 章では、利用者特性を考慮したサービスの詳細について述べる。4 章では、実際の利用を前提としたシステムの実装について述べる。最後に 5 章では、まとめと今後の課題について述べる。

2. 既存サービスの問題点と提案手法による解決法

2.1. 既存サービスの問題点

WWW の利用者の大部分がディレクトリサービスを利用しているにもかかわらず、そのほとんどがサービスに満足していない[4]という現状がある。これは、以下の二つの問題から生じていると考えられる。

(1) 情報への索引付け

従来のサービスでは管理者が情報一つ一つを閲覧して、その情報を表現する適切なタイトルやコメント、キーワードを手作業で付与しなければならなかった。キーワードの自動抽出はテキスト主体の情報には適用できるが、画像などのマルチメディア情報には適用することができない。さらに、管理者がその情報に対して詳しい知識を持っていないければ、適切なコメントやキーワードを付与することは困難である。そのため、索引として付与されるコメントやキーワードは管理者の主観に大きく依存することになり、情報を適切に表現しているとは必ずしも言えないのが現状である。

(2) 情報の分類

管理者が管理しやすい分類が利用者にとって利用しやすい分類とは限らず、多くの場合両者はトレードオフの関係にあるといえる。特に分類の基準の設定に、サービスの利用者の声がほとんど反映されておらず、索引付けと同様に管理者の主観に依存するところが大きい。また、分類に対して情報を追加・削除するにも、サービスの管理者が手作業で行う必要があり、サービスの構築・保守の面からも管理者への負担が大きいといえる。

以上の二つの問題が、サービスの管理者の負担を増加させ、結果的に利用者には不満を与えている主な原因であると考えられる。

2.2.提案手法による解決法

本研究では、サービスの利用者の興味や傾向をプロフィールという形式で表現[5]する。このプロフィールは、キーワードの集合であり、各々のキーワードは「単語」と「重み」、そして「期限」の組で表現される。単語と重みから、利用者がある情報に対してどれだけ興味を持っているか、どのような情報を好む傾向があるかなどを獲得することが可能となる。またプロフィールは利用者が情報を閲覧する度に更新されるため、サービスの利用が進むほど利用者の興味を的確に表現できるようになる。また、期限という概念を導入することで、利用者の興味の変化を追従することが可能となる。さらに、プロフィールを利用することで、サービスの利用者がどのような情報を必要としているのか、どのようなサービスを望んでいるのかなど、利用者要求の獲得やウェブマーケティングへの応用[5]も将来的には考えられる。

また、情報閲覧時に用いる索引をインデックスという形式で表現する。このインデックスは、情報のタイトルや URL などのメタ情報と、キーワードの集合から構成される。キーワードは単語と重みの組からなり、情報に対して形態素解析などの自然言語処理を行って獲得する。

本研究における提案手法の特長は、情報にアクセスした利用者のプロフィールをインデックスにフィードバックする点にある。初めに獲得したキーワードに、実際に情報を閲覧した利用者のプロフィールからキーワードがフィードバックされることにより、情報に含まれる文字としてのキーワードから、利用者の興味を踏まえて情報を適切に表現するキーワードへと昇華していく。本手法より、適切な索引付けにかかるサービスの管理者への負担を軽減することが可能となり、利用者もより使いやすい索引を利用することができる。また本手法では、画像などのマルチメディア情報に対しても順次キーワードが付与されていくため、テキスト主体の情報と同様に扱うことができるという利点がある。

3.利用者特性を考慮したサービスの提案

3.1.提案するサービスの概要

本研究で提案するディレクトリサービスとは、利用者の興味や傾向を管理者の索引付けや分類に利用することにより、管理者の負担を軽減し、利用者により使いやすいサービスを提供することを目的としたものである。そこで本研究では、以下の二点に着目する。

- (1)プロフィールの獲得
- (2)インデックスの更新

まず、利用者の興味や傾向を表現するプロフィールを導入する。更に、情報の内容を適切に表現するインデックスを作成して、このインデックスに対してプロフィールをフィードバックする手法を提案する。この手法により、従来管理者により行われてきた索引付けや分類にかかる諸問題の解決を図る。(図1)

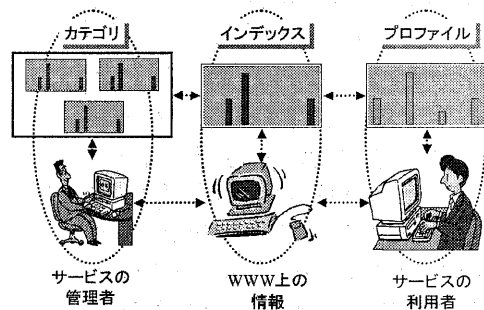


図1 提案するサービスの概要

また、本研究で提案するサービスは、大きく以下の三つの機構から構成される。

- (a)プロフィールの獲得機構
- (b)インデックスの作成機構
- (c)インデックスの更新機構
- (d)カテゴリの分類機構

本稿では、上記の(a)および(c)について述べている。(b)および(d)については現在詳細化を進めており、順次サービスに組み込んでいく予定である。

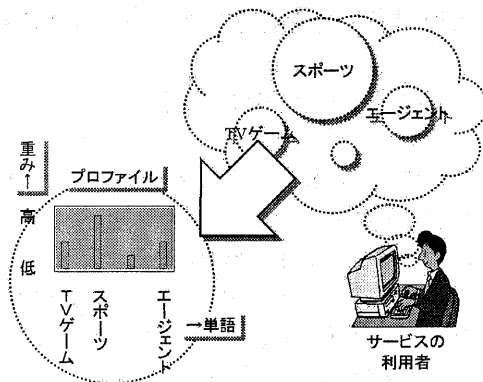


図2 プロファイルの概念図

3.2. プロファイルの獲得

プロフィールとは、サービスの利用者がどのような情報に対して興味を持っているか、またどのような情報を好む傾向があるかなど、利用者の特性を表現するものである。(図2)

プロフィールはキーワードの集合であり、各々のキーワードは単語と重み、それに期限の組で表現される。(式1)

$$\begin{aligned} Profile &:= \{ID, Keyword\} \\ ID &:= \{Name, E-mail, etc\} \\ Keyword &:= \langle Term, Weight, Limit \rangle \end{aligned}$$

式1 プロファイルの定義

○プロフィールの獲得について

サービスの利用者がある情報を閲覧したいと思い、後述するインデックスを選択したとする。利用者は通常のブラウジングと同様に目的の情報を閲覧するが、並列して以下の処理が行われている。まず、インデックスに含まれるキーワードの集合とプロフィールに含まれるキーワードの集合が抽出される。抽出されたキーワードを比較して、プロフィールに含まれるキーワードの重みの再計算を行い、次式のようにプロフィールを更新する。

$$K_profile' \leftarrow K_profile + \alpha \cdot K_index$$

ここでは、 $K_profile()$ は更新前(更新後)のプロファイルのキーワード、 K_index は更新前のインデックスのキーワードであり、 α はインデックスからの反映の度合いを決定する係数である。

3.3. インデックスの更新

インデックスとは、情報の内容をより適切に表現した単一の閲覧用索引である。(図3)

インデックスは、タイトルやコメントなどのメタ情報と、キーワードの集合から構成される。プロフィールと同様にキーワードは単語と重みの組で表現される。(式2)

$$\begin{aligned} Index &:= \{MI, Keyword\} \\ MI &:= \{Title, URL, Comment, etc\} \\ Keyword &:= \langle Term, Weight \rangle \end{aligned}$$

式2 インデックスの定義

○インデックスの作成について

初めに形態素解析を行い、TF・IDF法[6]を用いてキーワードの集合を作成する。管理者は実際に情報を閲覧して、情報を適切に表現するタイトルやコメントなどのメタ情報を手作業で付与する。付与されたメタ情報に、作成したキーワード集合を加えることによって、インデックスを作成する。

○インデックスの更新について

サービスの利用者の情報閲覧時に、サービスの管理者側にプロフィールが送信されてくる。このプロフィールの集合を一つの特性とみなして、キーワードの比較と重みの再計算からインデックスを更新する。

$$K_index' \leftarrow K_index + \beta \cdot K_profile'$$

ここでは、 K_index は更新後のインデックスのキーワード、 β はプロフィールからの反映の度合いである。

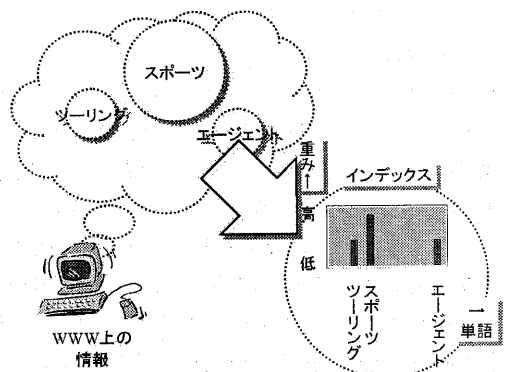


図3 インデックスの概念図

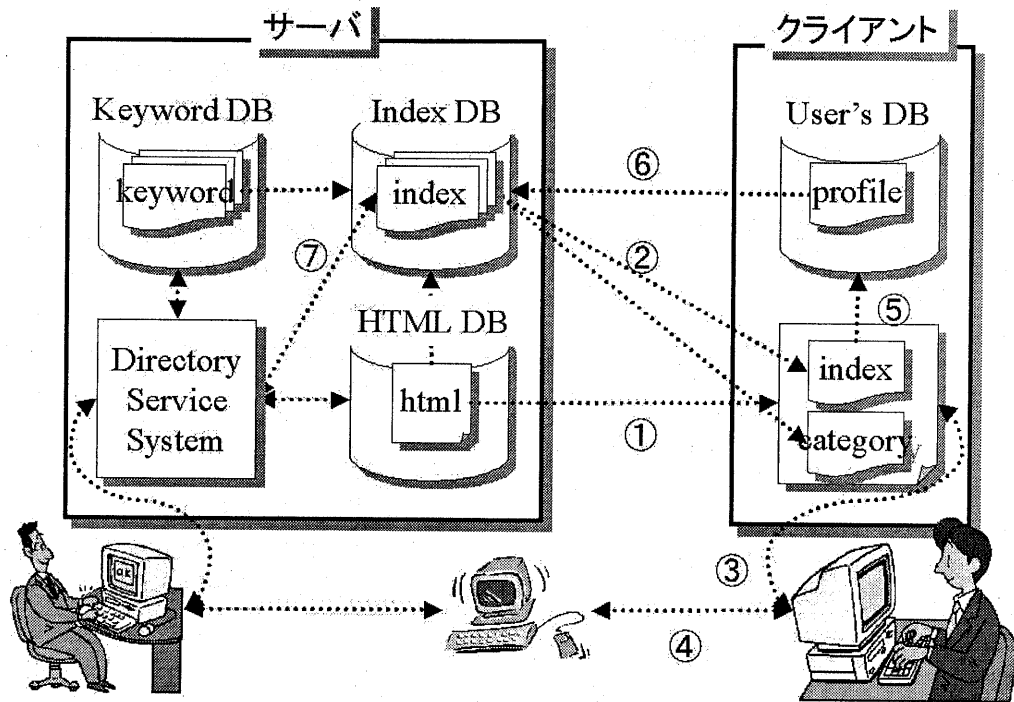


図4 実装システムの処理手順

4. 提案するサービスの実装

4.1. 実装システム

提案するサービスの機構のうち、今回はプロフィールの獲得機構とインデックスの更新機構を実装した。実装システムにおける処理手順は以下のようになる。(図4)

サービスの利用者が、ブラウザを通じてサービスにアクセスするところからスタートする。

- ① アクセスしたページに対応する HTML ファイルを、サーバからクライアントに送信する。
- ② さらに、①のページを構成するインデックスファイルをクライアントに送信する。
- ③ 送信されてきたインデックスの中から、利用者が興味を持った情報を選択する。
- ④ URL から、利用者は WWW 経由で③で興味を持った情報を閲覧する。

①から④までが、サービスの利用者による情報の閲覧である。この処理と並列して、プロフィールおよびインデックスの更新が実行される。

⑤profileの更新(クライアントで処理)

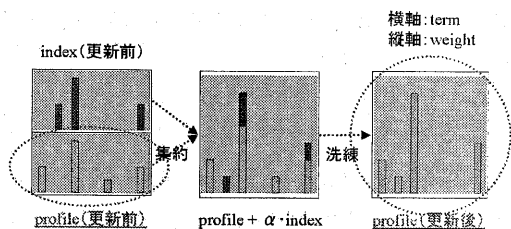


図5 プロファイルの更新

- ⑤ インデックスに含まれるキーワード集合をもとに、プロフィールを更新する。(図5)
- ⑥ 更新されたプロフィールがクライアントからサーバに送信される。

5. おわりに

本研究では、従来のディレクトリサービスにおける索引付けや分類の問題を解決することにより、管理者が管理しやすい、そして利用者が利用しやすいサービスを提供することを目的とした。そのために、利用者の興味や傾向を表現するプロフィールを導入し、これを情報に付与されるインデックスに対してフィードバックするという手法を用いることで、より適切なキーワードを持つ索引の作成を支援した。本稿では、プロフィールの獲得機構およびインデックスの更新機構までを実装し、提案手法の有効性を検討した。

今後の課題として、獲得したインデックスを用いた分類の半自動化を可能とするアルゴリズムの提案および情報分類機構の実装などがあげられる。

参考文献

- [1]「Yahoo! Japan」, <http://www.yahoo.co.jp/>
- [2]「Vector」, <http://www.vector.co.jp/>
- [3]「Yahoo!の歴史」
<http://www.yahoo.co.jp/docs/pr/history.html>
- [4]「第2回 FRI インターネットユーザ調査」
<http://www.fujitsu.co.jp/hypertext/fri/cyber/>
- [5]「An agent system for learning profiles in broadcasting application on the Internet」
C. Cuenca, LNCS1363, 1998
- [6]「WebMate : A Personal Agent for Browsing and Searching」Liren Chen, 1997
- [7]「パーソナライズ情報提供方式の提案と評価」
橋高, 情報処理学会論文誌 Vol.40 No.1 p175-

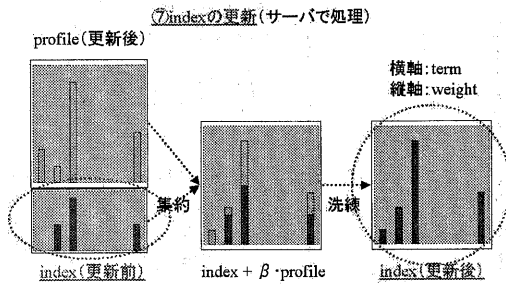


図 6 インデックスの更新

- ⑦複数のプロフィールに含まれるキーワード集合をもとに、インデックスを更新する。(図 6)

サービスの利用者が図 2 のプロフィールで表現され、なおかつ閲覧した情報が図 3 のインデックスで表現されたとする。このとき、更新後のプロフィールとインデックスはそれぞれ図 5 と図 6 になる。以上のように、本システムにより、利用者の特性や情報の内容を的確に表現することが可能となる。

4.2. 既存サービスとの比較

本研究で提案するサービスと「Yahoo! Janan」、それに NTT ソフトウェア研究所のパーソナライズ情報提供方式である「InfoBrocket」[7]について、比較を行った。(表 1)

◎…対応済, ○…一部対応, △…未対応

本研究のサービスは、索引を随時更新するために検索の精度は低くなるものの、特に情報への索引付けや分類の自動化に関しては、既存のサービスや研究に比べて有利であるといえる。

	Yahoo! Japan	InfoBrocket	本研究のサービス
利用者特性の考慮	△	◎	◎
索引付けの自動化	△	○	◎
分類の自動化	△	△	○
検索の精度	○	○	△
情報の推薦	△	◎	○

表 1 既存サービスとの比較