

# マルチエージェントに基づく WWW 情報検索支援

## WWW Information Retrieval Support Based on A Multi-agent System

山田晃弘<sup>\*1</sup>, 小久保卓<sup>\*2</sup>, 北村泰彦<sup>\*3</sup>

Teruhiro YAMADA, Takashi KOKUBO, Yasuhiko KITAMURA

<sup>\*1</sup> 財団法人イメージ情報科学研究所

Laboratories of Image Information Science and Technology

<sup>\*2</sup> 京都大学大学院情報学研究科

Graduate School of Informatics, Kyoto University

<sup>\*3</sup> 大阪市立大学工学部

Faculty of Engineering, Osaka City University

**Abstract:** Generally speaking, there are two ambiguities in WWW information sources and in user's query, when we retrieve WWW information by using a search engine. To deal with the ambiguities, we developed "keyword spice", a technology that agentizes a generic search engine into a domain specific information agent, and a multi-character interface technology for WWW information retrieval.

### 1. はじめに

今日 WWW (World Wide Web) は急速にわれわれの社会に浸透し、われわれの日常生活を支えるインフラストラクチャの一つとして、電子商取引、研究や教育活動、個人やグループによる情報発信、コミュニティの形成など様々な目的のために利用されている[1]。WWW に蓄積されている情報は日々、急速な勢いで増加を続けているが、その反面、必要な情報の検索はより困難なものとなっている。

WWW 情報検索の手段としては現在、検索エンジンが最も広く利用されている手段といえるが、例えば、豚肉のレシピ(料理法)を知りたい利用者が「豚肉」というキーワードを検索エンジンに入力した場合には、レシピだけでなく、畜産、小売、レストランなどを含む約 35,000 件のページがヒットしてしまい、また、必要な情報だけに絞り込むには経験や知識が必要である。このことは特に高齢者などの計算機の初心者にとって使いにくいものとなっている。

一般に検索エンジンによる WWW 情報検索には二つの曖昧さが存在するといえる。一つは WWW 情報源の曖昧さである。すなわち、現在の WWW ページは主に HTML により記述され、十分な意味構造をその中に記述することが難しい。したがって現在の検索エンジンでは入力さ

---

連絡先：  
山田晃弘 財団法人イメージ情報科学研究所  
〒559-0034  
京都市中京区河原町通二条下ル一之船入町 366  
E-mail: yamateru@kuis.kyoto-u.ac.jp

れたキーワードがページに含まれてさえいればヒットしてしまうという問題がある。もう一つは利用者の要求表現の曖昧さである。例えば、ある利用者が「豚肉」というキーワードを入力したとしても、その背景には「今日は豚肉を使った料理を作ろう。昨日は中華風だったから、今日は和風にしよう。少し健康にも気をつけたほうがいいからカロリーは控えめに。」といった要求が存在することも考えられる。もちろん現在の検索エンジンはこのような利用者の要求をそのまま受け付けることはできず、その要求が圧縮された少数のキーワードのみを受け付けることになる。

そこでわれわれは、WWW 情報源の曖昧さの問題に対し、汎用検索エンジンの出力をフィルタリングすることにより専門情報エージェント化する検索隠し味の技術を、利用者の要求表現の曖昧さの問題に対し、マルチキャラクタインタフェースの技術を開発した。以下では、この開発した2つの技術について述べる。

## 2. 検索隠し味

WWW 情報検索の手段として、現在最も広く利用されている検索エンジンではあるが、検索結果が膨大な量である場合が多く、また、その検索結果から必要な情報だけに絞り込むことも難しい場合が多い。

このため WWW からある特定の内容のページだけを検索する専門検索エンジンについても様々な研究が行われている。例えば利用者から与えられたページを解析して得られたキーワードを汎用検索エンジンにかける WebMate[2]や、利用者から与えられたページの関連ページを効率よく収集するアルゴリズムの研究も行われている[3]。また、汎用検索エンジンの結果をフィ

ルタリングし個人のホームページだけを検索する Ahoy![4]などがある。しかし、いずれも、高齢者などの計算機の初心者にとって十分に使いやすいものであるとはいえない。

われわれのアプローチは、以下の通りである。検索エンジンを利用する際の問題点は単純なキーワードを入力するとそのキーワードに関連する様々な領域の大量の WWW ページが出力されることである。その絞り込みにはキーワードの追加が一般的であるが、どのようなキーワードを追加すべきであるかは利用者の経験と知識に依存することが大きい。例えば、デジタルシティ京都[5][6]には京都に関するホームページが4500件以上登録されているが、京都の「観光案内」を検索したい場合には「おいでやす」をキーワードとして入力するのが有効である。また、豚肉のレシピに関して検索したい場合には、「豚肉 and 料理」をキーワードとして入力するよりも、「豚肉 and 塩」を入力したほうが有効である。

そこでわれわれは、このような WWW 情報検索の経験や知識からしか得られない追加キーワードを、利用者に代わって自動的に追加することで情報のフィルタリングを行う情報検索支援技術を開発した。ここで、利用者には見えない追加キーワードのことを、われわれは「検索隠し味」と呼んでいる。この「検索隠し味」を使うことで、汎用検索エンジンの出力をフィルタリングし特定の内容のページだけを検索でき、専門情報エージェント化することを可能にした。

「検索隠し味」を利用した WWW 情報検索は以下のような手順で行われる。

1. 自然言語による検索要求を受け付ける。  
「豚肉を使った料理を教えて」
2. 形態素解析によりキーワード抽出を行う。

- 「豚肉」、「料理」
- 不要キーワードを削除する。  
「豚肉」
  - 領域限定キーワードを追加する。  
「豚肉」、「塩」
  - 汎用検索エンジン(Goo など)で検索する。

これにより、WWW 情報検索の経験や知識のない利用者でも、WWW から効率よくレシピページだけを検索することができる。

次に、領域に特化した「検索隠し味」を見つけ出す手順について述べる。

- 領域に関連する一般キーワード(例:「ラーメン」、「焼肉」など)を用いて得た検索エンジンの出力結果を得る。この WWW ページの集合を S とする。
- 集合 S を領域に関連するもの(例:レストラン) T と、関連しないもの F に分類する。
- 集合 T に含まれる頻出キーワード(例:上位 50 個)を求める。これを「隠し味候補」と呼ぶ。
- それぞれの隠し味候補に対して以下の適合率を求める。  
適合率 = (T での出現ページ数) / (S での出現ページ数)
- この適合率が高いものが検索隠し味となる。
- ただし、適合率が低い場合は、隠し味候補の AND 結合を用い、より適合率の高いものが検索隠し味となる。

以上の手順で実際にレストラン検索における検索隠し味を求めた結果を表 1 と表 2 に示す。ここで、集合 S (全体) 5341 ページのうち、集

表 1. レストラン検索隠し味

キーワード	集合T(レストラン)での出現ページ数	集合F(レストラン以外)での出現ページ数	適合率
定休	471	338	0.582
住所	335	360	0.482
営業	676	743	0.476
駐車	262	362	0.420
:	:	:	:

表 2. レストラン検索隠し味(AND 結合)

キーワード	集合T(レストラン)での出現ページ数	集合F(レストラン以外)での出現ページ数	適合率
定休	471	338	0.582
定休&住所	239	115	0.675
定休&住所&営業	235	104	0.693
定休&住所&営業&駐車	49	41	0.544
:	:	:	:

合 T (レストラン) は 953 ページ、集合 F (レストラン以外) は 4388 ページであった。

表 1 からレストラン検索においては、「定休」が最も適合率が高い隠し味候補であるが、適合率が 0.582 と低いため、表 2 のように上位隠し味候補の AND 結合を用いた結果、「定休 and 住所 and 営業」が 0.693 と適合率が高くなる。よって、レストラン検索においては「定休 and 住所 and 営業」が検索隠し味となる。

さらに検索隠し味は再帰的に利用することもできる。例えば、ある検索隠し味を用いてレシピページのみを抜き出し、さらにそこから和食、洋食、中華などの検索隠し味を用いてレシピのジャンルを分類することも可能である。

実際に和食レシピ検索における検索隠し味を求めた結果を表 3 に示す。ここで、集合 S (レシピ) ページのうち、集合 T (和食) は 1684 ページ、集合 F (洋食及び中華) は 2674 (1844+830) ページであった。

表3. 和風レシピ検索隠し味

キーワード	集合T(和食)での出現ページ数	集合F(洋食・中華)での出現ページ数	適合率
煮出し	435	7	0.984
昆布	223	9	0.961
出し汁	393	30	0.929
みりん	685	57	0.923
:	:	:	:

表3から、和食レシピ検索においては、「煮出し」、「昆布」、「出し汁」、「みりん」が検索隠し味となる。また、同様に処理した結果、洋食レシピ検索においては、「生クリーム」、「ワイン」、「ブイヨン」、「バター」、中華レシピ検索においては、「中華」、「ごま油」、「土ショウガ」、「スープ」が各検索隠し味であった。

検索隠し味は情報検索の立場からは、再現率を犠牲にして、適合率を向上させる手法であるといえる。したがって再現率を向上させるためには、複数の検索隠し味を OR 結合して用いることが考えられる。例えば、豚肉を使った和食

料理であれば、「豚肉 and (材料 or 作り方) and (煮出し or 昆布 or 出し汁 or みりん)」などである。

### 3. マルチキャラクタインタフェース

キャラクタインタフェースは従来のキーボードや GUI よりもさらに利用者に優しいインタフェースとして期待されている[7]。これは基本的に、音声対話に基づく人間同士のコミュニケーションに近い、最も自然なインタフェースといえるからである。近年、マイクロソフトは MS-Agent の開発に力を入れており、WWW 上のインタフェースとしても、Extempo ( <http://www.extempo.com> )、Hapttek ( <http://www.hapttek.com> )、Virtual Personalities ( <http://www.vperson.com> )、Artificial Life ( <http://www.artificial-life.com> ) などのキャラクタインタフェースを利用した商用システムがすでに開発されている。また、キャラクタインタフェースによる対話エージェント

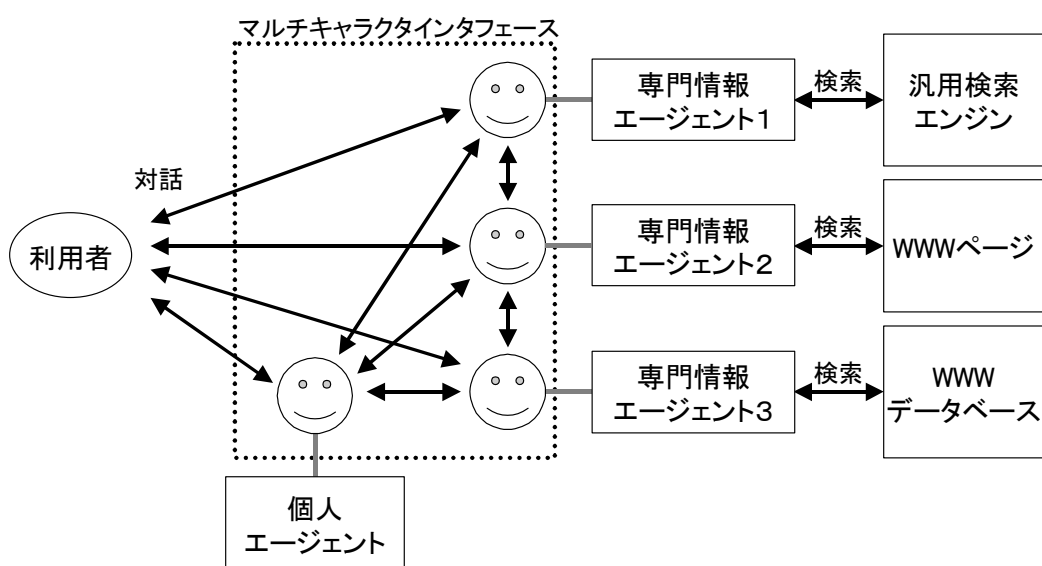


図1. マルチキャラクタインタフェースを用いた WWW 情報検索支援システム

トと情報検索エージェントに分離したマルチエージェントシステムの研究も行われている[8]。

しかし、これまでのキャラクタインタフェースは利用者とキャラクタの1対1の対話を対象とするものであった。そのため、利用者の要求の曖昧さに対応するには、自ずとキャラクタが細かく利用者に質問や確認する必要があり、利用者の対話回数も多くなっていた。

そこでわれわれはマルチキャラクタのインタフェースを開発した。これは、個人情報管理する個人エージェントキャラクタと、複数の専門情報エージェントキャラクタが、情報検索の過程をキャラクタ同士の対話で利用者に表現しながら WWW 情報検索を進めていく。図1にマルチキャラクタインタフェースを用いたシステムの構成図を示す。

図2に個人エージェントとレシピ専門情報エージェントの対話シーン例を示す。このシーンは、個人情報を知らないレシピ専門情報エージェントが、まず、一般的な料理レシピ(例：豚肉料理)を検索してくるが、そこで、利用者の情報を知っている個人エージェントが、昨日利

用者が食べたジャンル(例：和食)と違うジャンル(例：中華)での検索をレシピ専門情報エージェントに依頼している様子である。

このように、利用者の代わりに個人エージェントキャラクタが WWW 情報の検索絞り込みを行っていく。

ここで、キャラクタ同士の対話に利用者の要求と異なるものがあつた場合は、利用者はキャラクタ同士の対話に割り込み、情報検索の修正を要求することを可能としている。

これは、準知的なエージェントからなるマルチエージェントシステムにおいて、エージェントが利用者の要求の曖昧さを解消する上での自然な方法となる。例えば、次のようなシナリオを考えてみる。

1. 利用者 個人エージェント：「豚肉を使った料理を食べたい。」
2. 個人エージェント 利用者：「それではレシピエージェントを呼び出します。」
3. 個人エージェント レシピエージェント：「昨日は和食を推薦したので、今日



図2．マルチキャラクタインタフェース  
(レシピ専門エージェントが一般的な豚肉料理レシピを検索してきた際に、個人エージェントであるオウムが、利用者の情報から、昨日と違う中華での検索を依頼している様子)

は豚肉を使った洋食を推薦して下さい。」

4. 利用者 個人エージェント：「今日も和食が食べたいな。」
5. 個人エージェント レシピエージェント：「それでは豚肉を使った和食を推薦して下さい。」
6. レシピエージェント 利用者：「それではこんなのはどうでしょう。」
7. 利用者 レシピエージェント：「それはカロリーが高すぎるので、別のものにしてください。」

ここで個人エージェントは利用者の検索履歴を保持しており、それに応じてレシピのジャンルを切り替える機能があると仮定している。またレシピエージェントはカロリーに応じたレシピの推薦ができると仮定している。ただ問題は利用者の好み（例：和食好き、カロリーが気になる）をどのようにして知るかという点であるが、マルチキャラクタインタフェースではそれぞれのエージェントからの応答に対して利用者が個別に反応（割り込む）できることにより、情報検索のどの過程で、何が利用者の意図に反していたのかを的確に把握でき、それぞれのエージェントは利用者の好みを学習することができる。すなわち、利用者の要求の曖昧さを解消していくことができる。

このような好みの学習を単一のインタフェースエージェントで実現しようとする、そのエージェントは利用者の好みに関するあらゆるパラメータを処理できる能力が必要であり、その実現は難しい。

#### 4 . おわりに

WWW情報検索における WWW 情報源の曖昧さ及び、利用者の要求の曖昧さに対処するための、汎用検索エンジンの出力をフィルタリングすることにより専門情報エージェント化する検索隠し味の技術と、マルチキャラクタインタフェースによるWWW情報検索支援について述べた。

今後の課題としては、これらのエージェントを協調させることにより、より高度な情報検索や情報統合を実現することである。その一方、情報検索への利用者の要求が複雑になると、キャラクタエージェントの動作も複雑になることが予想される。利用者に対してはキャラクタのふるまいはできるだけ単純であることが望ましいので、それに関しては今後の課題である。

#### 5 . 謝辞

本研究は新エネルギー・産業技術総合開発機構 (NEDO) の委託事業「シニア支援システムの開発」として、高齢者のインターネット利用支援のために、イメージ情報科学研究所において、京都大学、大阪市立大学、三洋電機、NTT 西日本、NTT コムウェアの共同研究として行われたものである。特に、ご指導いただいた京都大学石田亨教授に感謝いたします。

#### 参考文献

1. 石田晴久：インターネットはここまで来た - 現状と課題 - , 情報処理, Vol.39, No.5, pp.393-399, 1998
2. L. Chen and K. Sycara, WebMate: A Personal Agent for Browsing and Searching, September 1997. <http://www.cs.cmu.edu/~softagents/webm>

- [ate/aa98-webmate.ps](#)
3. M. Hersovici, M. Jacovi, Y. S. Maarek, D. Pelleg, M. Shtalhim, and S. Ur, The shark-search algorithm. An application: tailored Web site mapping, Computer Networks and ISDN Systems. Proceedings of the 7th International World Wide Web Conference, Brisbane, Australia, vol. 30(1-7), pp. 317-326, April 1998.
  4. J. Shakes, M. Langheinrich, and O. Etzioni, Dynamic Reference Sifting: A Case Study in the Homepage Domain, 6th International World Wide Web Conference, Santa Clara, CA, April 1997.<http://ahoy.cs.washington.edu:6060/doc/paper.html>
  5. Toru Ishida, "Digital City Kyoto: Social Information Infrastructure for Everyday Life," Communications of the ACM (CACM), 2001 (to appear).
  6. Kaoru Hiramatsu, Kenji Kobayashi, Ben Benjamin, Toru Ishida, and Jun-ichi Akahani, "Map-based User Interface for Digital City Kyoto," The Internet Global Summit (INET2000),[http://www.isoc.org/inet2000/cdproceedings/4c/4c\\_1.htm](http://www.isoc.org/inet2000/cdproceedings/4c/4c_1.htm), 2000.
  7. Elisabeth Andre and Thomas Rist. Adding Life-Like Synthetic Characters to the Web. Matthias Klusch and Larry Kerschberg (Eds.) Cooperative Information Agents IV, Lecture Notes in Artificial Intelligence 1860, 1-13, Springer, 2000.
  8. Satoshi Oyama, Kaoru Hiramatsu and Toru Ishida, "Cooperative Information Agents for Digital Cities," International Journal of Cooperative Information Systems, 2000 (to appear).