

## 口コミ機構のコスト制御法

南 俊朗

minami@flab.fujitsu.co.jp

大谷 武

ohtani@flab.fujitsu.co.jp

(株) 富士通研究所 ネットメディア研究センター

〒814-8588 福岡市早良区百道浜 2-2-1 富士通九州 R & D センター

口コミ機構は、口コミエージェント間の局所的ネットワークを介して需要のある地域へ適応的に情報が伝えられるという性質を利用し、検索エンジンの持つトラフィック集中や情報の陳腐化等の問題解決を目指している。情報伝達は仮想的な伝達コストにより制御されるため、伝達コストの配分法は本機構にとって重要である。本稿では、近隣のエージェントに伝えられる情報に割り当てる伝達コストが一様の場合と可変の場合に、適応性の指標である平均アクセスコストがどのように異なるかを調べたシミュレーション実験について述べる。実験の結果、可変方式の方がより適応的であることが確かめられた。また、広告コストが適応性の違いに大きな影響を与えることが分かった。

## Controlling the Word-of-Mouth Mechanism with Distribution Costs

Toshiro Minami

Takeshi Ohtani

Netmedia Lab., Fujitsu Laboratories Ltd., Fukuoka 814-8588 Japan

The word-of-mouth mechanism intends to utilize its adaptability feature for solving the problems of the WWW search engines such as concentration of network traffics, latency of searched results, and so on. The adaptability that information is circulated towards the areas where the demands is high depends on the use of the virtual distribution-costs. This paper reports some results of the experiments for comparing two methods of cost-distribution; the uniform distribution and one in which cost varies according to the demands. The cost-varying method is more adaptive than the uniform-cost method in these experiments. We found also that the advertisement cost gives significant effects to adaptability.

### 1 はじめに

近年、World Wide Webを中心にインターネットへの情報発信が活発化し、ネットワークを通じて大量の情報が入手できるようになった。しかし、同時に、大量情報の中に欲しい情報がまぎれてしまい、必要な情報の発見が困難となった[3]。この問題に対処するため、検索エンジン[1, 2, 4, 5, 10]による情報資源探索支援サービスが提供され、我々は世界中に散在する非常に多くの情報資源の存在を知り、利用することが可能となった。しかし、検索エンジンにも、以下のような問題がある。

**資源情報の陳腐化：** 検索結果に基づき情報にアクセスすると、サービスを停止していたり、場所や内容が変わっていたりすることが意外に多い。資源情報の収集を頻繁に行うと、このような資源情報の陳腐化を防ぐことができるが、それでは、サー

バ負荷やトラフィックの増大を招くことになる。

**検索結果の精度問題：** 検索エンジンにより得られた結果も依然大量である場合、それらの中から有用性の高いものを選択する必要がある。しかし、多くの検索エンジンで、文書中に出現するキーワードの個数を有用性の基準としている。そのため、実際の有益さが反映されず、文字の少ない文書は低く評価される反面、有用そうなキーワードが数多く現れる文書は高く評価されてしまう。

**負荷の増大とトラフィックの増加：** 検索エンジンは資源情報を集中管理していることが多いため、利用者のアクセスの集中により、検索エンジンの負荷の増大やレスポンスの悪化を招きがちである。また、Lycos[4]やWorld Wide Web Worm[5]のようにロボットによる資源情報の自動収集も、端

から機械的に文書を漁るため、無用トラフィックや Web サーバの負荷の増大を招く。

このような問題に対処する1つの方式として我々は、情報の伝達範囲が限定されており、また、様々な評価情報が行き交うという口コミ的な特徴をもった情報流通の仕組みとして、“口コミエージェント”の連携により、ネットワークにおける情報収集と収集した情報資源の分散管理を行う情報資源の探索システム（口コミ機構）[7, 8, 9]の研究・開発を行っている。本機構は(1) 資源情報の分散管理により、ネットワークの環境変化への適応性が高い、(2) 情報伝達が局所的で、有用情報が需要の高い地域に伝搬されるために、トラフィック総量が減少し、レスポンスも向上する、(3) 大量の情報からの有効情報の選択に役立つ評価情報を伴った資源情報の流通が組み込まれている、(4) 情報資源へのアクセス状況に関する評価情報の自動収集により、評価が速やかに利用される、等の特徴を用いて、上記の問題の解決を図っている。

我々は、口コミ機構のこれらの特徴の内、特に需要に応じた情報分布を誘導する適応性に関するシミュレーション実験を行った。本実験を通じて、コストを周りの口コミエージェントに平等に伝える方式とコストの配分を過去の検索要求の実績の応じてコストの配分を可変とする方式を比較・検討し、より優れたコスト制御法の開発を目指している。以下、第2節では、口コミ機構の基本的な動作や特徴を概観し、第3節で、シミュレーション実験の方法ならびにその結果を示し、考察を加える。最後に第4章で、全体のまとめ及び今後取り組むべき課題について述べる。

## 2 口コミ機構の基本的動作

口コミ機構は、情報資源の広告と検索を行う口コミエージェントのネットワークで構成される。各口コミエージェントは、資源情報提供者から、情報資源に関する広告情報を受け付け、自分の資源情報データベースに登録する。同時に、近隣の口コミエージェントにも広告情報を伝える。この時、口コミシステムの管理者が定める広告伝搬に用いる仮想的な“コスト”によって資源情報が伝達される範囲が制限される。コストとしては、最初に情報提供者から広告を受け付けた時点からの経過時間や、各口コミエージェントの利用や通信にかかる費用とも、仮想的な“費用”とも定めることができる。広告が口コミエージェント間を中継されるにつれて、与えられたコストは消費され、それがなくなると伝搬は終了する(図1右半分)。

一方、検索を行う利用者も最寄りの口コミエー

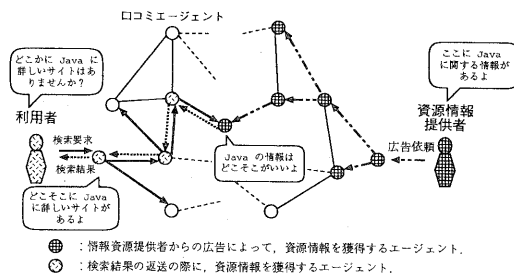


図 1: 資源情報の広告と検索

ジェントに検索要求を出す。それを受けた口コミエージェントは、自分の資源情報データベースを検索し、該当するデータがあれば、それを利用者に返送する。なければ、情報資源提供者の広告と同様、近隣の口コミエージェントに検索要求を伝える。ここでも、広告の場合と同様に、検索要求のための“コスト”により伝達範囲が限定される。広告情報を持つエージェントへ検索要求が到達すると、検索成功となる。検索結果は、検索要求が配送されてきた経路を逆に辿り、返送される。その際、各口コミエージェントは、送られて来た検索結果の内容を調べ、その中に自分の資源情報データベースにない情報、あるいは、より新しい情報が含まれていたならば、自分の資源情報データベースを更新する。このようにして、情報提供者が最初に広告を行った範囲外にも広告情報が伝達される(図1左半分)。すなわち、多くの利用者が検索を行うことで、資源情報がより広範囲に伝達され、資源情報を有する口コミエージェントの数が増加する。情報が検索要求の多い地域へと伝達され、要求のない地域には伝わりにくいという適応性を持つ。そのため、最初の検索で満足のいく結果が得られなくても、次の検索では有益な結果を得ることが起る。これは、誰も参照しない資源情報を無闇に持ち続けて、計算機資源を無駄にすることが少ないという利点にもつながる。

## 3 シミュレーション実験

### 3.1 実験の概要

**ネットワーク結合** 図2に示すように400個の口コミエージェントを、縦横20の格子状に結合した口コミネットワークを構成する。一辺の長さ $S(=20)$ をサイズと呼ぶ。更に上辺と下辺、右辺と左辺のそれぞれをつなぎトラス状にすることで、端をなくし、全てのエージェントを同一状況とする。エージェント間を移動するために辿る結合の数の最小値をそれらの距離と定める。すなわち、座標

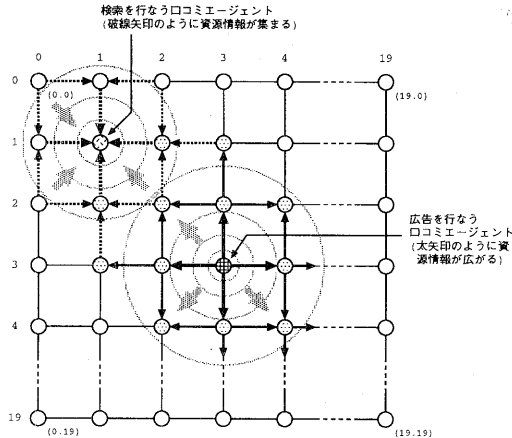


図 2: 20×20 の格子状口コミネットワーク

$(i_0, j_0)$  のエージェントと座標  $(i_1, j_1)$  のエージェント間の距離は

$$\min(|i_1 - i_0|, S - |i_1 - i_0|) + \min(|j_1 - j_0|, S - |j_1 - j_0|)$$

である。これより、トーラスの直径 (エージェント間の距離の最大値) は、縦横のサイズ  $S$  が偶数のときは  $S$ 、奇数のときは  $S-1$  となり、本例の場合、20 である。

**広告** まず、ランダムに選ばれた広告情報受け付けエージェントは、上下左右の4方角へ、与えられた広告コストを用いて広告情報を伝える。広告情報を受け取った各エージェントは、情報元の方角を除く3方角のエージェントへ、受け取った広告コストを元に計算された広告コストを付与して情報を転送する。各伝達の際、コストは1だけ消費されるものとする。従って、隣のエージェントが情報を受け取ったときにその情報の広告コストは送った時のコストよりも1だけ少なくなる。広告情報に対して最初与えられた広告コストは消費され減少していき、伝達コストが1より小さくなると、そこで伝達は終了する。

情報に与えられたコストを周りのエージェントへ平等に割り振る場合、距離  $n$  ( $n > 0$ ) 以内のエージェントの全てに情報を伝達するために必要な最低コスト  $c$  は、 $6n(n-1) + 4$  となる。例えば、 $n = 1$  (4つの隣接エージェント) の場合必要なコストは4となる。各エージェントにコスト1が割り当てられ、伝達後のコストは0となり、その先へは伝達されない。 $n = 2$  の場合のコストは、16となる。これが4つの隣接エージェントに4ずつ割り当てられる。情報が転送された結果、それぞれの隣接エージェントは、コスト3で情報を受け取る。受け取った情報は、情報を渡した広告源

のエージェント以外の3方角のエージェントへ1ずつ分配され、さらに、それらに情報が送られる。情報はコスト0で受け取られるため、それ以上の転送は行われない。このようにして、情報は距離2以内のエージェントへ伝えられる。以下、 $c$  を伝達コスト、 $n$  を距離コストと呼ぶ。実際のコスト計算は伝達コストを用いて行い、表示等の際は、実際の距離と対応した距離コストを用いる。

**検索** 広告が行われた後、繰り返し仮想的な検索を行う。検索コストは広告コストとは別途与えられる。距離コスト  $n$  で検索を行うと、距離  $n$  以内に広告情報を持ったエージェントが1つでも存在するならば、検索に成功し、発見された情報は両者をつなぐ途中のエージェントを経由して検索要求を発行したエージェントに伝えられる。その際、途中のエージェントにも広告情報が記録される。

**試行及び効果の指標** 実験毎に500回の検索要求を発生させ、情報検索を行う。適応性の指標として、各エージェント毎に情報へ到達できる最小コストを距離コストに換算し、全てのエージェントに渡る最小距離コストの平均値 (平均アクセスコスト) および情報を持つエージェント数が試行5回毎に計算される。

**検索発生への偏り** 検索の偏りによる適応性の違いを見るために、検索要求の発生が一様で偏りのないパターン (Bias=0)、標準分布パターン (Bias=1)、そして、偏りを強調したパターン (Bias=3) での結果を比較する (図3)。後の2パターンは、座標が (10,10) のエージェントで最も検索発生確率が高く、そこから遠ざかるにつれて発生確率が低くなる。

**コストの配分** 2つのコスト配分方式に関してシミュレーション実験を行い、両者を比較する。一方は、各エージェントに平等にコストを割り振る一様方式である。他方は、過去の検索要求の履歴に応じて、検索要求を伝える際の検索コストに差をつけるコスト可変方式である。具体的には、各エージェントは過去どの方角のエージェントから何件の検索要求があったかを記録しておく。検索の要求を行う頻度の大きい方角のエージェントは情報を知っている可能性が高いという仮定の下で、届いた検索要求全体に対する各方角からの検索要求の割合に応じてコストを割り当てる。すなわち、過去の検索要求の多い方角のエージェントへの検索コストを大きく、少ない方角のエージェントへは検索コストを小さく割り当てる。初期状態では一様方式と同様とする。

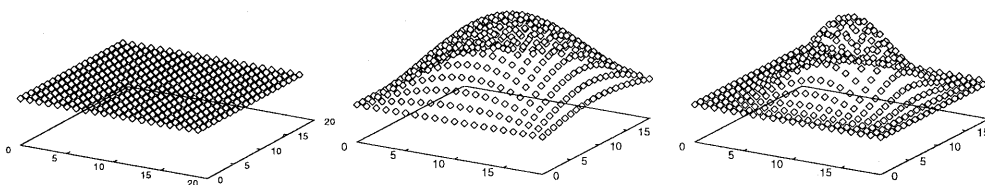


図 3: 検索分布関数の形状 (左: Bias=0, 中: Bias=1, 右: Bias=3)

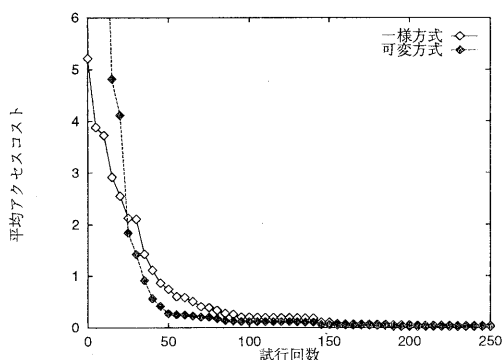


図 4: 平均アクセスコストの変化

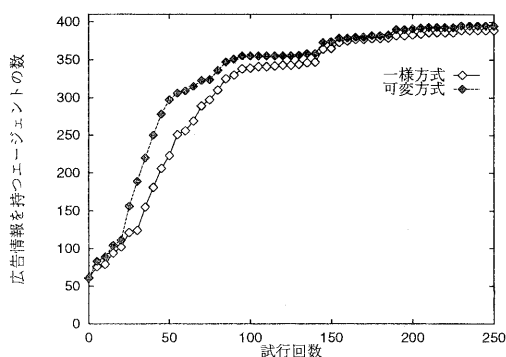


図 5: エージェント数の変化

### 3.2 実験結果

広告、検索の両コストとも5で、標準パターンでの検索発生の場合の平均アクセスコストの変化を図4に、広告情報を持つエージェント数の変化を図5に示す。一様方式と比べ可変方式のアクセスコストは、最初非常に大きく、試行が進むにつれて急激に減少する。20～30試行以降可変方式のコストは一様方式より小さくなる。エージェント数に関しては初めから可変方式の方が大きい。

状況をいろいろと変化させ、様々な実験を行った結果、次の知見が得られた。すなわち、一般に広告情報を持つエージェント数が大きければ、検索の際、近隣に情報を持つエージェントが存在する可能性が高まり、その結果平均アクセスコストは小さくなる。従って、情報を持ったエージェントの分布が需要に適應しているかどうかは、単に平均アクセスコストを比較するのではなく、エージェント数が同じ場合での平均アクセスコストで比較するのが妥当である。

エージェント数に対する平均アクセスコストの比較を図6に示す。この図より、エージェント数が全体の半分程度以下の場合、可変方式による平均アクセスコストは一様方式より大きい。それより大きくなると、可変方式がわずかに有利となっている。この事実より、エージェント数が半数を越

えた領域で、コスト可変方式の平均アクセスコストが一様方式のコストより勝っていたのは、実はエージェント数が勝っていたことが原因であったことになる。

次に、口コミ機構での情報伝達に関連したパラメタ値の違いによる、平均アクセスコストの違いを調べる。図7に広告コストを、図8には検索コストを、そして、図9に、検索発生への偏りを変えた場合の、エージェント数に対する平均アクセスコストの変化を示す。これらの図より、検索コストや検索の偏りの違いによる平均アクセスコストの変化の差はあまりないことが分かる。実際、これらの違いは、同じパラメタ値でのばらつきと同程度であり、誤差内の違いであると言える。それに対して、広告コストの違いによる平均検索コストの差異は明らかである。

この原因として考えられるのは、本実験では最初に一度だけ、与えられた広告コスト内の距離にあるエージェントに広告情報を伝える点である。すなわち、広告の伝達に関しては最初は需要への適應が行われておらず、検索が発生することで、はじめて、検索需要に応じた情報伝達が行われ、情報を持つエージェントの数の増加が起る。大きな広告コストの広告と小さなコストの広告が行われ、その後、本方式の適應的なやり方で情報が伝えられた場合を比較すると、同じエージェント数

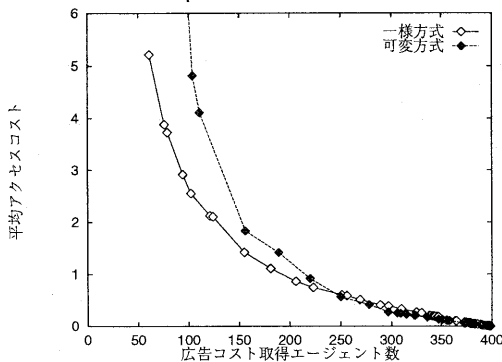


図 6: エージェント数に対する平均アクセスコスト

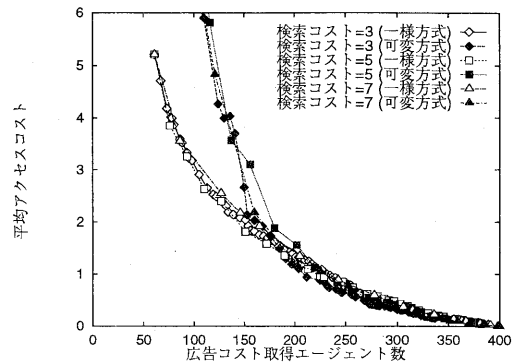


図 8: 検索コストによる違い

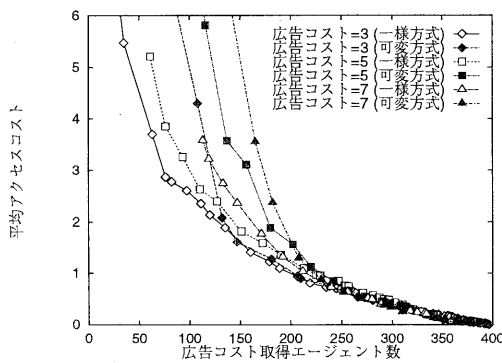


図 7: 広告コストによる違い

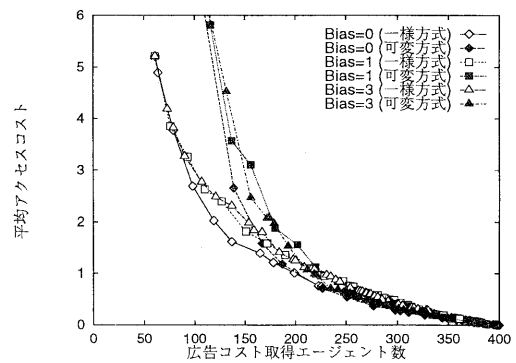


図 9: 検索の偏りによる違い

で比べた場合、前者では、検索需要に対する適応があまりなされないでその数に達する。後者は広告コストが少ないうえ、その後の検索、すなわち、適応が多く行われ、その結果規定のエージェント数に到達する。その結果、広告コストの小さい方が、より環境に適応し、平均アクセスコストも小さくなるのである。

一方、検索コストは、広告コストと異なり、検索が起ってはじめて意味をもつ。口コミ機構においては、検索の際、コストの大小による多少の違いはあるにしても、基本的には、検索が成功してはじめて、検索エージェントへと広告情報が伝えられる。検索コストが小さいならば、検索に失敗する可能性は大きい、その代わり、失敗時には情報を持つエージェント数の増加も起らない。結局、検索コストが大きいならば、検索に成功したとき、情報を持つエージェント数の増加が大きく、速やかに情報が広まる点で異なるのみで、最終的に規定のエージェント数に達した場合の適応の度合いは、検索コストの大小に寄らず同程度であるということになる。

検索分布の偏りによる違いの理由も検索コストと同様である。すなわち、分布の偏りが大きい時、広告情報は検索が多発する領域に速やかに伝えられる反面、その領域内のエージェントのみに情報が保持される。その結果、検索の偏りにより、エージェント数の増加は抑えられ、平均アクセスコストの減少も抑えられてしまう。このことは、口コミ機構の持つ適応性自体が潜在的に持つ問題点を示している。すなわち、口コミ機構は適応的であるがために、検索の少ない領域のエージェントには情報が存在せず、そこで発生した検索要求は失敗する可能性が大きいことになり、検索の少ない領域からの検索に対する必要コストが大きい。強すぎる適応性を、抑える仕組みが必要である。

#### 4 まとめ

口コミ機構は、ネットワークを構成したエージェントの、周囲との局所的な情報伝達を通じて情報の広告や検索を行う分散型情報検索機構である。情報が伝えられる範囲は、仮想的に設定された、

伝達のためのコストによって限定される。情報への需要に応じて徐々に情報が広まることで、口コミ機構は需要に対する自然な適応性を持つ。

本稿は、伝達コストを隣接するエージェントに平等に割り振る一様方式と過去の検索要求の履歴に応じて割り振る検索コストを変化させる可変方式のコスト管理法をシミュレーション実験により比較することで、口コミ機構の適応性の違いを調べた結果、可変方式は一様方式よりも急速に（広告情報を持つ）エージェント数が増加し、従って適応性が高いことが確かめられた。また、適応性の指標とした平均アクセスコストはエージェント数との関連が高いこと、すなわち、（1）エージェント数が全体の約4割以下ならば、エージェント数に対する平均アクセスコストは一様方式の方が少なく、可変方式ではエージェント数が少ないほど、急激に平均アクセスコストが増加すること、（2）それを越えると、一様方式と可変方式の両者の差がほとんどなくなることで、分かった。

従って、エージェント数が少ないとき、一様方式の方が平均アクセスコストは小さい、しかし、可変方式の方がエージェント数の増加が急速であるため、平均アクセスコストの減少も急速であり、数十回の試行により一様方式を追い越す。

また、情報伝達のためのパラメタの違いによる平均アクセスコストの比較の結果、広告コストの違いは平均アクセスコストの違いをもたらすこと、そしてこれは、口コミ機構の持つ適応性に起因していること、また、検索コストや検索の発生分布の違いは、機構が元々もっている適応性が働き、方式の違いが生じないことが分かった。

現在、口コミエージェント機構のプロトタイプシステムを開発し、それを用いた実験を行っている。今後口コミエージェント数、データ両、そしてユーザ数を増やした実験を行い実システムでの検証を進める。また以下の課題にも取り組む。

- 実験を更に進め、コスト配分方式についてより精密な評価を行うとともに、それを踏まえ、過適応を抑えるなどコスト配分方式を改良する。
- 口コミ機構は、情報を保持するエージェント数を少なく抑えつつ低コストでの検索の実現を狙っている。そのため、コスト配分法の工夫だけではなく、それほど重要性の高くない情報に関しては、近隣のエージェントがその情報を持つならば、自己の情報を削除するといった、より緻密な情報管理方式の開発が求められる。
- 現在、口コミエージェントの近隣関係は、管理者があらかじめ設定することとしている。資源情報をやりとりする内に、近隣にない口コミエージェントの存在を知り、その情報を利用して、

口コミエージェントのネットワークを自動的に更新し、パフォーマンスを向上させる仕組みも有用性が高い。

- 我々は、[6]において、ネットワーク雑誌帳ZKの、言語的データと非言語的データの両者を含んだハイブリッドデータの流通・利用への口コミ機構の応用例を示した。ユーザがそれぞれの人的ネットワークを通じ行うコミュニケーションの支援に口コミ機構を適用することで、文字通りの口コミネットワーク支援システムを実現することができる。このような有益なアプリケーションシステムの開発にも力を注ぎたい。

## 参考文献

- [1] AltaVista: <http://altavista.digital.com/>
- [2] P.M.E. de Bra and R.D.J. Post: Information Retrieval in the World-Wide Web: Making Client-based searching feasible. In O. Nierstrasz, editor, *Proc. of the First Annual World Wide Web Conference*, 1994.
- [3] 服部文夫: ネットワークエージェントによる情報収集と流通. 学会誌「情報処理」, Vol. 38, No. 1, pp. 30-35, 1997.
- [4] M.L. Mauldin and J.R.R. Leavitt: Web Agent Related Research at the Center for Machine Translation. In *Proc. of the ACM Special Interest Group on Networked Information Discovery and Retrieval (SIGNIDR-94)*, 1994. <http://fuzine.mt.cs.cmu.edu/mlm/signidr94.html>.
- [5] O.A. McBryan: GENVL and WWW: Tools for Taming the Web. In *Proc. of the Fifth International World Wide Web Conference*, pp. 15-29, 1994. <http://www.cs.colorado.edu/home/mcbryan/mypapers/www94.ps>.
- [6] T. Minami, H. Sazuka, S. Hirokawa, and T. Ohtani: Living with ZK—An Approach towards the Communication with Analogue Messages. In *Second International Conference on Knowledge-based Intelligent Electronic Systems (KES'98)*. IEEE South Australia Section, 1998.
- [7] 大谷武, 南俊朗: 口コミによる情報資源探索. 第6回マルチ・エージェントと協調計算ワークショップ, 1997. <http://www.kecl.ntt.co.jp/msrg/macc97/ohtani.html>.
- [8] 大谷武, 南俊朗: エージェント社会における資源管理. 情報処理学会 研究報告, 98-ICS-110, pp. 55-62, 1998.
- [9] T. Ohtani and T. Minami: The Word-of-Mouth Agent System for Finding Useful Web Documents. In *Proc. of the Asian Pacific Web Conference (APWeb98)*, 1998. (to appear).
- [10] Yahoo: <http://www.yahoo.co.jp/> や <http://www.yahoo.com/>