

## 3T-02 Web 情報収集エージェント Mobeet の評価

森川 景介† 深澤 良彰† 吉岡 信和‡ 大須賀 昭彦‡ 本位田 真一††  
 †早稲田大学理工学部 ‡(株)東芝 ††国立情報学研究所

### 1. はじめに

インターネットの急速な発展とその上での情報の充実によりインターネットは、様々な疑問に回答を出してくれる巨大なDBになってきている。この状況下では、必要な情報を効率よく探す方法が重要になる。

このような背景で Web 情報をユーザの代りに探してくれるエージェント(Mobeet[1])が開発された。本稿では、エージェントを利用することで情報の収集をどれだけ効率化できるのかについて、実験データを基に考察を行なう。

### 2. Mobeet とは

この実験で使用する Mobeet(モビート)とは、マルチエージェントフレームワーク(Bee-gent[2])をベースとした検索エージェントである。

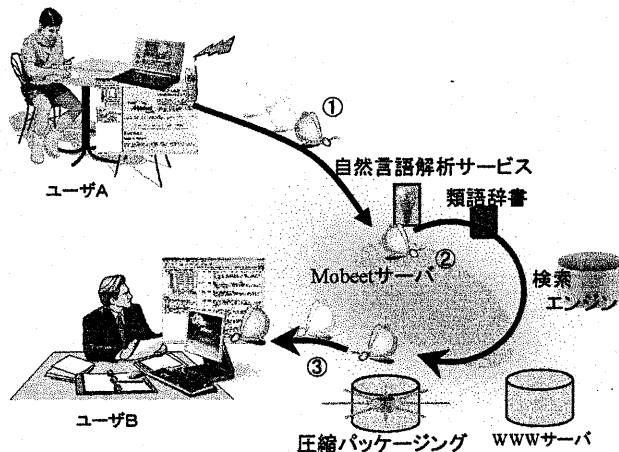


図1: Mobeet の仕組み

Mobeet の仕組みは、まず①ユーザAはエージェントへの指示を自然言語で行い、サーバに送信する。②エージェントはユーザの代理としてサーバ上で Web の情報検索を続行する。サーバ上では入力されたキーワードを言語解析し、そこで解析された情報を検索エンジンで検索する。そして集めてきたページのコンテ

ツを解析し、フィルタリングする。これらの作業が終了するとエージェントサーバで待機し、③ユーザBは結果を都合の良い時、都合の良い場所で受け取る。エージェントは、ユーザからの帰還要求を受け取ると、コンテンツを圧縮して帰還する。ユーザはエージェントが収集した情報をオフラインで通信コストを気にせず見ることができる。

Mobeet は以下のような3つの特徴を備えている。

- 高度な検索
- 移動エージェントによるまとめ取り
- 自然言語入力

本研究では、以上の特徴を各々評価していくこととする。

### 3. 高度な検索の実験・評価

Mobeet が集めてきた情報は、人間がコンテンツを直接検索するのに比べてどの程度有用かを調べた。

まず、Mobeet にとって集めやすい情報、集めにくい情報という特性があるのかを知るために、一般的にどのようなニーズ(何を Web 検索に取得してほしいか)があるかを分析した。そのために実際の公開サービスの入力ログデータを入手・分析し、それについてカテゴリ分けした。その中から、いくつかの検索キーワードを考え、Mobeet、一般的な検索エンジン(以下: 検索エンジン)を使用し情報検索をした。

実験方法は、まず Mobeet を使用しエージェントに指示を与え、ネットワークに送る。そして収集作業が終わり、エージェントがネットワークからユーザの環境上に帰還した時点から、ユーザが回線をオフラインにして、キャッシュに収められている情報を見終わるまでの時間を計測した。一方、一般的な検索エンジンでもそれと同時間だけ情報収集を行い、それぞれの実験について得られた情報の質を5段階評価した。評価基準は表1に示す通りである。

表1: 評価基準

1	期待した情報が全く得られなかった
2	期待した情報がほとんど得られなかった
3	ほぼ期待した情報が得られた
4	期待した情報は全て得られた
5	期待以上の情報が得られた

Evaluation of Web Search Agent Mobeet.

Keisuke Morikawa †, Yoshiaki Fukazawa †, Nobukazu Yoshioka ‡, Akihiko Ohsuga ‡, and Shinichi Honiden ††.

†Department of Information and Computer Science, Waseda University

‡Toshiba Corporation

††National Institute of Informatics

806 件の実際の入力ログデータを分析し、カテゴリ分けをした結果は図2のようになった。

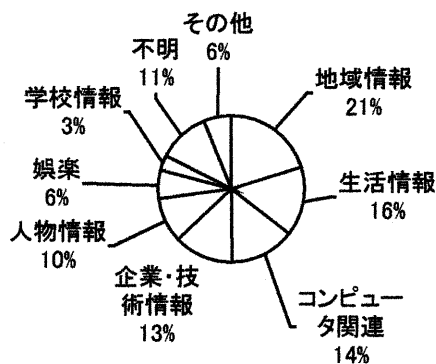


図2: 入力ログデータの内訳

この中の上位5位までのカテゴリの中から実験キーワードを用意した。

表2: 実験結果

実験NO	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Mobeet	3	4	1	2	1	4	3	4	3	4	3	4	3	4
検索エンジン	1	3	1	1	1	2	2	1	3	1	2	4	2	2
評価差	2	1	0	1	0	2	1	3	0	3	1	0	1	2

評価差=Mobeet の評価-検索エンジンの評価

表2の実験NO.1~5、6~8、9~11、12・13、14が上位5つのカテゴリに対応する。実験の評価差はランダムであることから、Mobeet にとってカテゴリによる情報収集能力の差はないと言える。

評価差を見ると、Mobeet を使ってエージェントに情報収集させた場合の方が、人間がコンテンツを直接探すのに比べて短時間で情報が収集できていることも分かる。よってエージェントが持ってくるコンテンツは、限られているが比較的適切にフィルタリングして選んできていると言える。

#### 4. 移動エージェントによるまとめ取りの実験・評価

エージェントにより Mobeet がどれだけ効率よくまとめ取りを行なっているかを評価する為に、サーバに帰還命令を送信してからユーザの手元に戻ってくるまでの通信量、通信時間、さらにオフラインにして Mobeet がまとめ取りしてきた URL(トップページ 10 件分)を 10 件分全て表示させるのに要した時間を計測した。同じ検索を検索エンジンで行い、キーワードを入力し検索ボタンを押した時点からの通信量、通信時間を測った。なお、この時検索する HP の内容は Mobeet、検索エンジン共全く同様のページで比較し、

通信量や時間の減少量を調べた。

表3: 実験結果

	平均送信バイト	平均所要時間	減少量
Mobeet	97883	1分23秒	送信バイト 25.59%
検索エンジン	131551	1分53秒	所要時間 26.55%

減少量=(1-Mobeet 平均/検索エンジン平均)×100

表3のような結果から、エージェントでまとめ取りすることで、オンラインで人間が直接検索するのに比べ通信量、時間共に約 25%減少することが分かる。これによりコスト面でのエージェント情報収集の優位性が分かる。

#### 5. 自然言語による検索の評価

実験を通して、自然言語から必要なキーワードの切り出しは適切に行なわれていることが分かった。しかし、形容詞などの修飾語、条件文の処理の不都合が生じることも分かった。

入力ログをみると“初めて”や“うまい”などの修飾語が使われている場合がある。現在 Mobeet は、キーワードとして名詞のみを切り出してくる。しかし、時に修飾語の有無によって検索内容が大きく異なる場合がある。そこで、単純な名詞の切り出しだけではなく自然言語の意味解析も必要である。

条件文の処理については、現行では“~以上”、“~以下”という条件で検索すると、単純にその文字列が含まれるコンテンツを集めてしまい、ユーザの要件を満たすことができない。解決には、自然言語の意味解析に加えコンテンツの意味解析も必要となる。

#### 6. まとめ

本稿ではエージェント情報検索についての高度な検索、移動エージェントによるまとめ取り、自然言語入力の実験を通して評価した。この結果、エージェントが Web 検索に有効であることが示された。

#### 参考文献

- [1] 情報収集エージェント: Mobeet  
<http://beegent.jaist.ac.jp:8000/Mobeet/>
- [2] マルチエージェントフレームワーク: Bee-gent  
[http://www2.toshiba.co.jp/beegent/index\\_j.htm](http://www2.toshiba.co.jp/beegent/index_j.htm)
- [3] 吉岡信和: マルチエージェントフレームワーク Bee-gent による Web 検索システムの開発, 情報処理学会第 60 回全国大会論文集, 1999(3)