

広域高速ネットワーク応用次世代インテリジェント分散システムの研究 (第2報)

エージェントによる生産支援技術の開発

村河亮利, 打田澄雄, 檜垣和生, 吉野信行

Study on advanced database search system with the Internet (2nd report) Development for manufacturing support technology with agent

MURAKAWA Akitoshi, UCHIDA Sumio HIGAKI Kazuo and Yoshino Nobuyuki

The period shortening is being obtained the productivity in the development initial stage of marketing research, planning, design, trial manufacturing and so on, since the unknown and trial-and-error factor is abounding. In this study, the technology by the fact for productivity improvement in the development stage is established with the Internet which is a rich information source. And demonstration by network on distribution system is carried out using the agent technology. Data collection is carried out for the specific home page as a part of the system this time, and the usefulness of this system is confirmed. And the followings are improved:

1. It adapted to the data collection of image processor, and the result was obtained on problem and practicability in respect of this system.
2. The operation from the browser became possible, and the operability of this system was improved.

キーワード: エージェント, ネットワーク, インテリジェント

1 緒 言

企業は顧客の望むものを必要なときに必要なだけ適正価格で供給できる生産体制を整えるため、さまざまな努力をしている。しかし、市場調査、企画、設計、試作などの開発初期における生産性は、未知、試行錯誤的要因が多いため、開発期間がボトルネックとなっている。

本研究では、共同研究公設試験研究機関および県内共同研究企業の協力を得て、豊富な情報源となったインターネットおよび高機能コンピュータを駆使することにより製品開発過程における生産性向上を図る技術を確立する。また、当センターのインターネットへの接続回線は 512kbps であるが、今後高速の通信インフラが整備されることを考慮に入れ、広島県が全県的な情報通信基盤として整備した高速ネットワーク（広島メイプルネット）を用いて本システムの実証実験を行う。同時に、製品・サービスへの適用を図ることで新市場の開拓を支援する。

また、電子回路設計支援のためのデータベース（以下 DB）を構築し、エージェント技術を用いて広域ネットワーク、分散システム上での実証実験を行う。今回は、電子回路設計支援の一部である画像処理装置の

効率的なデータ収集への適用を図った。

2 シ ス テ ム

2.1 システム概要

現在、分散 DB から目的とするデータを収集するシステム¹⁾²⁾や検索エンジン³⁾が開発されている。しかし、時々刻々と更新されるインターネット上のデータを収集し、使用者が望む情報を整理し提示するシステムは見当たらない。そこで、昨年度作成したデータ収集システム⁵⁾を使用して、現在インターネット上で広く使用されている Hyper Text Markup Language（以下 HTML）文書中から目的のデータを抽出・加工し、使用者に提示する手法について検討した。具体的には、まず使用者が必要な条件を入力することにより、あらかじめ使用者が登録したアドレス（以下 URL）、あるいはキーワードにより、検索エンジンから引き出したアドレスおよび独自に作成した DB から情報を引き出す。その引き出した情報から使用者が必要とするデータと一致した情報のみを提示することにより、使用者の作業負担を軽減し、効率よくデータ収集が可能なシステムとした。今年度は、昨年度のシステムを改良し、画像処理装置のデータ収集を行い、システムの検証お

よび問題点の抽出を行った。

2.2 システム構成

全体のシステム構成を図1に示す。

利用者は Web ブラウザを通して検索を行う対象物のキーワード、要求項目等を画面の指示に従い対話的質問に答える形式で入力を行う。入力されたデータは Web サーバを通じて検索サーバに送信され、インターネットや知識データベース、あるいは両方のデータベースから情報検索を行い、利用者が要求するデータを収集する。収集したデータから、利用者の要求項目とデータが一致するものを選択し、Web サーバを通じて利用者に提示する。

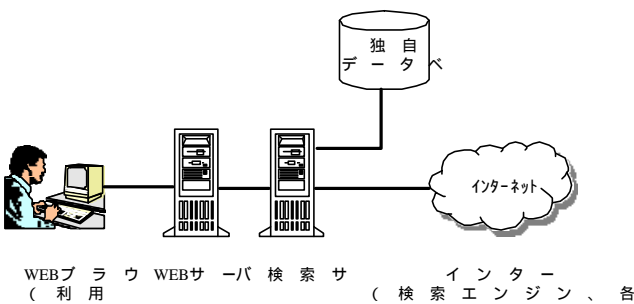


図1 システム構成

2.3 検索エンジンとの連携

利用者が Web 画面から入力したキーワードを基に自動的に検索エンジンにキーワード入力を行い、その検索結果を得る。その検索結果からリンク先の抽出を行い、このデータを基に各リンク先アドレス HTML データの自動収集を可能とした。

ある検索エンジンに画像処理装置に関するキーワードとして『画像処理』を入力したときの結果から、リンク先データのみ抽出した例を図2に示す。

<http://www.XXX.or.jp/>
<http://www2.YYY.or.jp/~aaa/labo.html>
<http://AAA.ZZZ.or.jp/~bbb>

図2 検索エンジンからのリンク先抽出結果例

2.4 複雑な表形式への対応

今回、ホームページは HTML で、かつ表形式で書かれているもののみを対象とした。各リンク先のホームページを自動収集させた結果、各ホームページの検査対象項目の様子は単純な表形式ではなく、入れ子の表になっている場合が多いことが判明した。これは、ホームページの文字や図の位置を制作者の意図する様

に表示するためと見られる。

そこで、入れ子の表になっても各表毎にデータを抽出するアルゴリズムを開発した。図3に入れ子表の HTML 記述例、図4に入れ子表の表示例を示す。また、開発したアルゴリズムを適用した出力結果を図5に示す。

```
<table border=1>
<tr><td>1</td><td>2</td><td>3</td></tr>
<tr><td> <table border=1>
<tr><td>10</td><td>11</td><td>12</td></tr>
>
<tr><td>20</td><td>21</td><td>22</td></tr>
>
<tr><td>30</td><td>31</td><td>32</td></tr>
>
</table>
</td></tr>
```

図3 入れ子表の例

		10	11	12
		20	21	22
		30	31	32
		--END		
	1	2	3	
	4	5	6	
	--END			

図4 入れ子表表示例 図5 入れ子表の抽出例

2.5 ホームページからのデータの抽出結果

画像処理装置の様子が記載されたホームページの検索結果の例を図6に示す。このホームページから表のデータを抽出した結果を図7に示す。図7に示すように、表はタブ区切りのテキスト形式になっている。

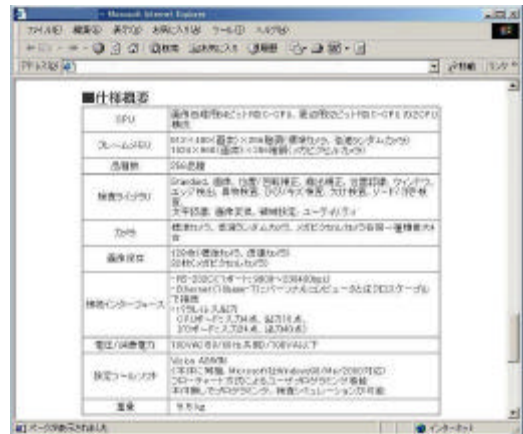


図6 画像処理装置検索結果例(一部)

```

http://www.XXX.co.jp/YYY/ZZZ/AAA.htm
2002/XX/XX 16:02:23
2002/YY/YY 10:49:15
CPU 画像処理用64ビットRISC-CPU 周辺用32ビットRISC-CPU の
2 CPU 構成
フレームメモリ 512×480 (画素)×256 階調 (標準カメラ, 倍速
ランダムカメラ) 1024×960 (画素)×256 階調 (メガピクセルカメラ)
:
カメラ 標準カメラ, 倍速ランダムカメラ, メガピクセルカメラ各同一種
類最大 4 台
:
電圧/消費電力 100VAC(50/60Hz 共用)/100 V A 以下
:
重量 5.5 kg
-----END OF http://www.XXX.co.jp/YYY/ZZZ/AAA.htm

```

図7 画像処理装置検索データ抽出結果 (一部)

2.6 キーワード抽出アルゴリズム

『画像処理』というキーワードで予め検索エンジンからの情報を基に7サイトのホームページから収集した約8,800ページを対象に,昨年度開発したキーワード抽出アルゴリズムを適用し,キーワードとその頻度について解析を行った。そこで『画像処理』の場合のキーワード抽出アルゴリズムの実用性について検証を行った。

2.7 知識データベース

知識データベースは,タブ区切りのテキスト形式となっており,2.5で述べた図7と同様の形式となっている。内容は,入力対象データの「項目」,項目に対する「データ」が記述される。最初の行にはタイトル名,2行目は空欄,3行目はデータ入力日であるが,ホームページから収集した場合はデータ収集日となる。2行目は,ホームページから自動収集を行った場合にホームページ作成日 (last-modified)が記入されるため,このデータフォーマットと一致させる必要があり,知識データベースでは空欄となっている。図8にその例を示す。

```

IP2000
2002/03/01
画像処理機能 1ドキュメントによる特徴量抽出
画像構成 1740*1024 画素
入力信号 1チャンネル濃淡映像信号
出力信号 1チャンネル濃淡映像信号, 1チャンネル2値化映像信号
2値化レベル 256段階(8ビット)
色抽出2値化方式
サブリクワット 72MHz
アスペクト比 1:1 (正方形画素)
計測項目 面積, 幅, 重心, 慣性等価楕円半径 (長軸長, 短軸長, 傾き角)
計測速度 ビデオレート(1/30秒)
インターフェイス PC98拡張バス
ボードサイズ 169*149mm
ドライブ C言語
-----END OF IP2000

```

図8 知識データベースの形式

3 Webブラウザによる操作

様々な端末から利用できるように Web ブラウザから操作可能とした。図9に初期画面, 図10にパラメータ等の設定画面を示す。

図9の画面から画像処理装置の例で言えば,プリント基板の画像処理装置を選び質問に答えながらホームページのリンクを辿ると,最終画面(図10)で,データが入力されている。これらの質問ホームページは使用者が事前に作成し,質問のページリンクを張っておく。

また,設定項目について知識がある場合は直接最終画面に飛び,直接項目の入力も可能である。

同義語,要求項目,キーワード等は予め初期値が設定されている。項目の登録,変更値の設定等を行い検索する。

図11に入力設定項目を示す。この項目を基に図8に示した知識データベースから抽出した出力結果を図12に示す。



図9 入力ブラウザ初期画面



図10 画像処理装置設定画面

キーワード 画像処理装置
項目 ルーム#:1024 インターフェイス:232

(1) IP:2000 2002/03/01 ルーム#:1024 インターフェイス:
(2) http://www.XXX.co.jp/YYYY/ZZZ/AAA.htm 2002/XX/XX 16:02:23 2002/YY/YY 10:49:15 ルーム#:1024 インターフェイス:232

図11 入力設定項目

図12 出力結果例

4 考 察

4.1 同義語

2.6で述べたキーワードについて、頻度が少ない単語については同義語としてまとめられるものが多く存在することが確認された。例えば、『画像転送先, 画像転送』や『レンズ, レンズオプション』などの2つ以上の語が連結したもののや、『センサ, センサー』などの同じ意味で表記が少し異なるものが多数確認された。これらの語を同義語として登録する必要がある。

この登録については、各分野独自に使用されるものであり、各分野に同義語としての辞書を設定する必要があると考えられる。

また、昨年度作成した同義語処理のアルゴリズムでは『センサ, センサー』を1語として処理可能であるが、『画像, 画像オプション』の様に漢字とカタカナが連結したものを1語としての処理が不可能であることが判明した。

4.2 操作性

Webブラウザから検索操作が可能となった。そのため、各端末にプログラムを導入する必要がなく、またブラウザはインターフェースとして広く使用されているため、操作性が向上した。また、3でも述べたが、質問については必要に応じて最後にパラメータの変更をすればよく、予め質問事項やパラメータを設定し初心者でもすぐに使用可能である。

4.3 知識データベース

今回作成した知識データベースはテキスト形式で、タブ区切りとした。インターネットから収集するデー

タとの形式を共通化することにより、プログラムから収集データと同様に扱うことが可能となった。

今後 既存のデータベースにも対応可能となるよう、SQLなどと連携を図る予定である。

5 結 言

利用者の要求するデータに適合するものを選択的に表示可能となった。またキーワードを入力し、検索エンジンから結果を抽出し、リンク先の整理を行うことが出来、以下のことが可能となった。

複雑なテーブルへの対応

ホームページからテーブルデータを抽出し、データベース化

知識データベースの検討および整理を行い、プログラムへの実装

画像処理システム用に調整を行った前処理部と後処理部を開発

今回は文字列の一致のみでのデータマッチングを行った。今後、数字や単位等を考慮し数量の大小を比較可能とする予定である。

また、各機能の抽出を行い、それぞれの検索エンジンやパラメータはファイル等の設定を変更を行い、プログラム本体の変更を少なくし、汎用性の高いシステムとする予定である。

文 献

- 1) 西野, 赤坂, 小泉: 情報処理学会論文誌 40(1999)1, 253.
- 2) 西野, 木室, 赤坂: 情報処理学会研究報告 98(1998)8, 201.
- 3) 森川, 深澤, 吉岡, 大須賀, 本位田: 情報処理学会講演論文集 3(2000), 367.
- 4) 茶筌: <http://chasen.aist-nara.ac.jp/index.html>
- 5) 村河, 打田, 中塩: 広島県立西部工技研究報告, No44(2001), 24