

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1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, 中塩武之

Study on advanced database search system with the Internet (1st report) Development for manufacturing support technology with agent

MURAKAWA Akitoshi, UCHIDA Sumio and NAKASHIO Takeyuki

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:

The time is shortened and the workload of user is reduced by automatic retrieval, the work which information retrieval and arrangement by hands. For example, it takes about 3 minutes by automatic arrangement function of this technique, on the other hand it takes about 30 minutes by hands.

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性能コンピュータを駆使することにより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64kbps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、最終的に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・サービスへ応用することで新市場の開拓を目指す。

本研究では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。今回そのシステムの一部として特定のホームページのみを対象にデータ収集を行い、本システムの有用性の確認を行った。

2 シ ス テ ム

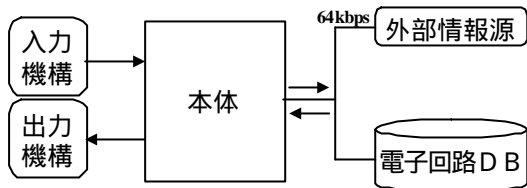
2.1 システム概要

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あたらない。そこで、このデータ収集を行うシステムを参考に、現在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）や独自に作成したDBから最新の情報を引き出す。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を行うシステムを開発した。

2.2 システム構成

システム構成図を図1に示す。構成として、入力機構、出力機構、本体、DBおよび外部情報源（インターネット等）で構成される。入力機構として、マイク、キーボードが接続されている。出力機構としてはスピー

ーカ、ディスプレイが接続されている。また、外部情報源の接続回線は 64kbps でインターネットに接続されている。



入力機構 :マイク、キーボード等
出力機構 :スピーカ、ディスプレイ等

図1 システム構成図

システム本体の外観を写真1に示す。



写真1 システム本体外観写真

2.3 システム仕様

今回用いた開発環境の仕様を表1に示す。

表1 システム仕様

項目	仕様
CPU	Intel Pentium (800MHz)
主記憶	128MB
外部記憶	30GB
NIC	10 Base T/ 100 Base TX
OS	Microsoft Windows 2000
開発言語	Microsoft Visual Basic
その他	音声認識ソフト,音声合成ソフト

2.4 ソフトウェア構成

ソフトウェア構成概略図を図2に示す。
初期設定は、ソフトウェア起動時にシステム設定等

の様々な初期化を行う。

登録設定は、使用者が情報検索を行うための要求項目、同義語定義、URL登録等を行う。

情報収集は、使用者が設定した項目を基に必要な情報源から収集を行い、中間DBに格納する。

情報分析は、中間DBから目的とする情報を分析抽出し、レポート作成を行う。

情報管理は、加工後のデータの編集、保存、削除、を行う。

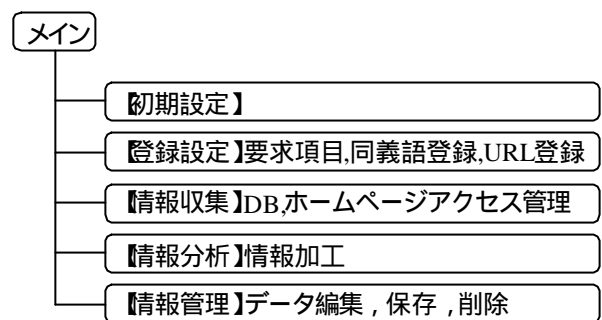


図2 ソフトウェア構成概略図

2.5 システム処理手順

本システムの処理の流れを図3に示す。まず使用者は音声によるマイク入力か、キーボードからデータ入力を行う。マイク入力の場合は、音声認識処理により文字列に変換される。その後、必要なDBにアクセスを行い、取り出した情報からキーワードの抽出を行う。切り出したキーワードから同義語の整理を行い、あらかじめ作ってある項目にそってデータの保存を行う。その後、使用者が入力したデータと一覧表を作成したデータとの文字列のパターンマッチングを行い、該当データが検出できた場合には、その検出データをディスプレイ表示または音声合成によって音声で使用者に示す。

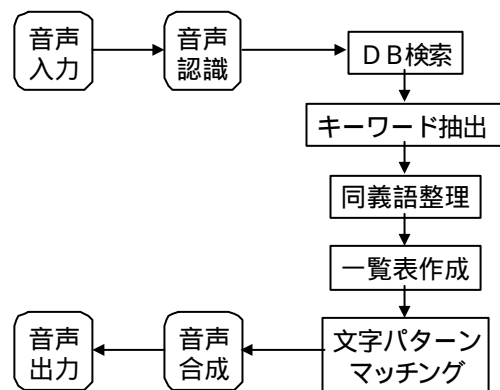


図3 システム流れ図

3 一覧表作成

3.1 一覧表作成概念

HTML 文書中のテーブルから一覧表作成までの流れを図4に示す。

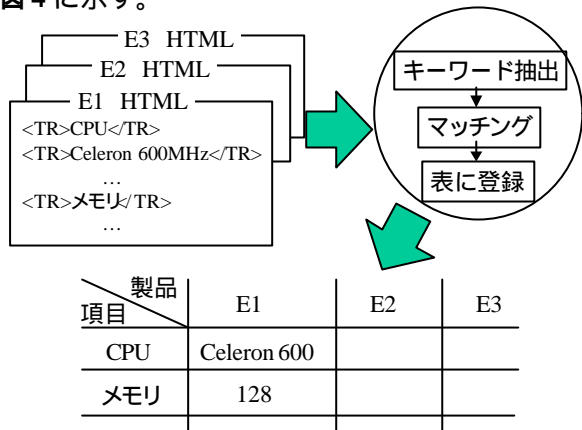


図4 一覧表作成行程

図4の表の空欄記入方法についての手順を示す。表の一番左側の列に使用者が要求する項目を記入。(例) 項目の1列目に中央処理演算装置(以下 CPU)、メモリと記入

HTML 文書からテーブル1行ごとにデータを抽出。

(例) <TR>CPU</TR><TR>Celeron 600MHz</TR>

そのデータをキーワードに分解。

(例) CPU,Celeron,600,MHz に分解。

で分解したキーワードを、項目とあらかじめ登録した項目に関係づけたデータと比較。

(例) のキーワード(CPU,Celeron,600,MHz)と、あらかじめ項目とデータとの関連を記載したデータ((例) CPU,Celeron,600)と比較。

の比較で一致すれば、表の該当箇所にキーワードを記入。一致しなければそのデータは破棄。

(例) CPUの項目の行にCeleron,600と記入。

同様の作業を、対象とする全てのHTML文書について実行。

登録作業終了後エージェントで判断を行うためのDBとして保存。

また、検索エンジンと組み合わせることにより、使用者が目的とするホームページのURLを自動的に検索することが可能となる。そのために検索エンジンの結果から目的のアドレスデータを切り出し、検索先アドレスデータとしてDBに保存する。

3.2 キーワード抽出方法

本システムでは、3.1で述べたようにデータから項

目に対するキーワードを抽出する必要がある。キーワードを抽出するために形態用語解析として茶筌³⁾がよく知られているが、本システムでは茶筌のように日本語を解析することが目的ではない。そのため、高速にキーワードを抽出する簡易な手法を考案し実証実験を行った。その状態遷移図を図5に示す。

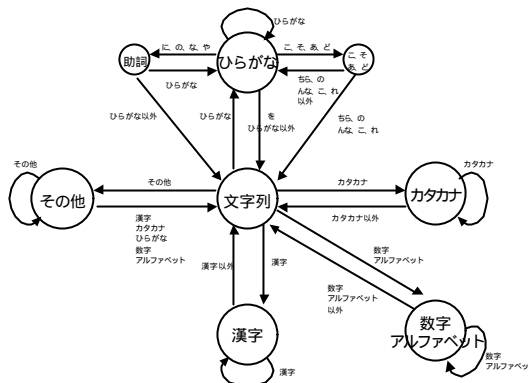


図5 キーワード抽出方法状態遷移図

まず、ある文字列の1文字目から検索を行う。1文字目がひらがなであれば、ひらがなに状態遷移を行う。2文字目が助詞の場合、助詞に状態遷移を行う。ここで3文字目がひらがな以外であればここが文字列の区切りとし、その区切られた文字列をキーワードとする。しかし、3文字目がひらがなの場合、再びひらがなに状態遷移を行い、文字列の区切りの検索を行う。同様にカタカナ、漢字、その他の文字の場合も行う。この状態遷移図を使用して今回キーワード抽出を行った。

3.3 一覧表作成(検索エンジンの場合)

検索エンジンから一覧表を作成するためのフローチャートを図6に示す。

使用者が質問の入力。

検索エンジンに適した検索コマンドを自動作成。

検索エンジンにアクセスを行い、検索結果を得る。

その結果データを基に、リンク先URL、キーワードおよび要約を抽出。

エージェント機能で判断させるためのDBとして抽出データを保存。

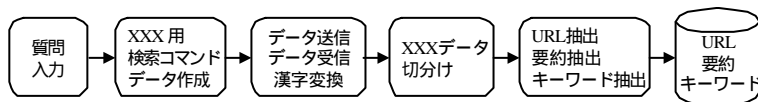


図6 一覧表作成(検索エンジンの場合)

4 操作

4.1 操作手順

コンピュータ仕様情報の収集・検索を例に操作手順

を示す。

必要な要求項目データおよび要求項目と関係するデータを書き込む(例: CPU, Celeron, 600, 800, 1)。ここで設定された情報に従って、目的の要求項目とキーワードの一致したものが目的とする表に書き込まれる。

同義語(例: CPU, プロセッサ, 中央演算処理装置)および検索を行う URL(例: http://xxx.yyy.jp)の登録を行う。登録画面を写真2に示す。

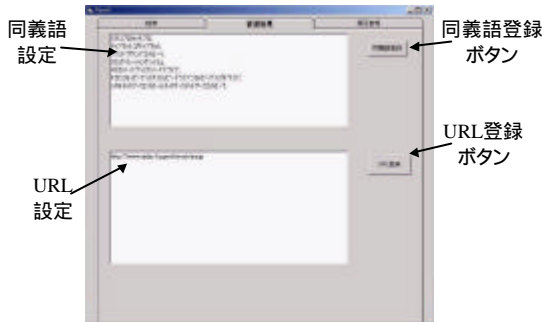


写真2 同義語およびURL設定画面

検索ボタンを押すと検索を始め、結果が表示される。なお同義語登録、項目登録については、一度登録すれば変更がない限り入力する必要はない。登録がすべて終了しているときは、要求項目のみを変更すればよい。結果および途中経過表示画面の様子を写真3に示す。

基本操作は音声入力でも同様な操作が行え、検索結果もディスプレイ表示だけでなく音声で出力可能である。



写真3 結果および途中経過表示画面

4.2 出力形式

収集したメモリ上のデータをファイルに保存し、表形式で表示した例の一部分を図7に示す。これは、コンピュータ仕様情報の収集・検索例である。先頭行は項目名、2行目以降は仕様データである。要求項目お

よび、仕様データはタブで区切られている。出力結果の2行目の例では、CPUとして800MHz, 1GHzあり、メモリは128MB, 256MBあり、チップセットは810を搭載していることを示している。

CPU	メモリ	FDD	チップセット
800,1,Celeron,Pentium,	RAM,128,256,	3.5,3,2,モード,	810,
Celeron,700,	RAM,64,128,256,	3.5,2,モード,	
Pentium,	RAM,128,256,1024,	3.5,2,モード,	

図7 データ出力結果例(一部)

5 考察

本システムを用いて特定のホームページ上からサンプルデータのデータ収集時間の計測を行った。表作成に必要なページは9ページあった。本システムを用いると約190秒程で目的の表が完成した。一方人間の手でホームページを開いて表にまとめる作業を行うと、30分以上要した。このことから、必要なホームページの数が増えるほど人手で行うには手数がかかり、さらにホームページ自体も不定期に更新されるため最新情報を得るには労力と時間がかかることがわかる。また、独自に開発したキーワード抽出アルゴリズムは、要求項目に一致するデータを振り分ける時に有効であることがわかった。

6 結 言

特定のホームページのみを対象にデータ収集を行い、時間短縮の面において本システムの有用性を確認した。また、最新のデータを収集することに当システムの有用性があることが示され、独自のキーワードアルゴリズムも有用であることが確認できた。今後、様々なホームページ上から収集したデータから、このシステムに適応した使用者が要求するデータのみを効率よく提示するための手法およびシステムの改良を行う予定である。

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.ja>