

情報利用の高度化のための知識ウェアハウスの提案とその実装

西尾 裕三、松尾 利行

株式会社日本総合研究所 アドバンスドテクノロジーグループ

〒550 大阪市西区新町1-33-8 井上特殊鋼本社ビル

TEL:(06)536-7631 FAX:(06)532-9701

{nishio,tosiy-m}@osa.sci.jri.co.jp

あらまし 情報の共有／再利用による企業活動の効率化は、生き残りのための最重要課題の1つである。我々は、情報利用の高度化の観点(情報収集・情報蓄積・情報共有)から知識ウェアハウスによる解決方法を提案し、情報処理に関する技術情報を対象とするプロトタイプを構築した。今後は、自己組織化された情報ベースの開発技術・知的エージェント技術・ドキュメントマイニング技術を適用し、最新情報の自動ダイレクトメールや最新のQ&A情報を使ったPCヘルプデスクのような実用システムを開発する予定である。

キーワード 知識ウェアハウス、情報収集、情報蓄積、情報共有、自己組織化、知的エージェント、ドキュメントマイニング

Proposal of a Knowledge Warehouse for advanced utilization of information and its development

Yuzo NISHIO and Toshiyuki MATSUO

Advanced Technology Group, The Japan Research Institute

ISS Bldg., Shinmachi 1-3-8, Nishi-ku, Osaka, 550 Japan

Tel:+81 6 536 7631 Fax:+81 6 532 9701

{nishio,tosiy-m}@osa.sci.jri.co.jp

Abstract Knowledge reuse/sharing is one of the most important subjects for enterprises to keep operating. We propose a Knowledge Warehouse in a viewpoint of advanced utilization of information(information gathering, information storage and information sharing). We have implemented a prototype system whose object is an information technology. We will implement a practical system such as a news agent or PC help desk to apply three technologies 1)self-organized information base, 2)intelligent agent and 3)document mining.

key words knowledge warehouse, information gathering, information storage, information sharing, self-organized information base, intelligent agent, document mining

1. はじめに

イントラネットが企業内で急速に普及し、情報システムの構成が大きく変化しつつある現在、情報の共有／再利用による効率化は企業の生き残りのための最重要課題の1つであり、そのためのしくみが渴望されている。この要件に対しグループウェアやワークフロー等の統合基盤システムが提供されているが、我々は情報利用の高度化の観点から知識ウェアハウスによる解決方法を提案する。

以下、2章では、我々の課題である情報処理に関する技術情報を例とし、共有／再利用の点から問題点を述べる。3章では、情報の高度利用の観点から知識ウェアハウスを提案する。4章では、知識ウェアハウスを実装するための基盤技術について詳細に述べる。5章では、情報処理に関する技術情報を対象として開発した知識ウェアハウスのプロトタイプシステムと、開発予定のシステム概要を述べる。6章では、まとめを行う。

2. 情報処理に関する技術情報

情報処理に関する技術情報を例として、その特質、共有と再利用の点からの問題点を述べる。

2. 1 特質

(1) 大量かつ未成熟な技術

他の業界と違い、情報処理業界で扱うべき技術が広範囲・多種類であり、かつ未成熟である。特に、この傾向は、メインフレーム中心の時代が終わり、オープン化の時代の到来に伴い顕著になっている。ハードウェア／ソフトウェアベンダーから新製品が提供される都度、技術者は、その製品の特徴を把握し、目の前にあるシステム開発案件への適用可能性を検討しなければならない。

(2) ノウハウの獲得

実際に使いこなした経験がある技術者だけが知っているその製品の癖（特性）もある。例えば、ベンダーが提示する能力（例えば、スループットなど）は、特性に最も精通しているベンダー自身がその特性を極限まで生かした結果であり、数字が示す通りの能力を引き出すのに経験が必要と

なる。ツールとツールの組み合わせについては、カタログ上の仕様を基に実際組み合わせず稼働できず、製品のベンダーに問い合わせをしカスタマイズした結果、稼働したという経験を持った技術者もいると思われる。ツールの組み合わせとハードウェアとの相性を考えるとその組み合わせの数はおそらく数千のオーダーになり、それらすべてを事前に確認することは不可能である。

2. 2 共有／再利用の面から見た問題点

情報技術は前節のような特質を持っているため、他の技術者が習得した情報を共有／再利用することが渴望されているにもかかわらず、プロジェクトチーム内・グループ内・メーリンググループ内のようなかなり限られた範囲（コミュニティ）内でのみ情報が共有されている場合がある。技術者の置かれている立場（広く技術を習得すべきジェネラリストあるいは個々の技術を深く精通すべきスペシャリスト）に応じて、情報技術の共有／再利用を提供する環境を整備しなければならない。

3. 知識ウェアハウスの提案

前節で述べた問題点は情報処理技術だけでなく、一般の技術にも当てはまると考えられる。我々は、情報利用の高度化の観点すなわち情報収集・情報蓄積・情報共有化の観点から、解決方法として知識ウェアハウスを提案する。

3. 1 知識ウェアハウスの定義

「ネットワーク上に散在する統制されていない情報を整理・統合することにより、迅速かつ確かな意志決定、生産性向上、新たな知識の創造などの企業活動を支援するための情報利用高度化システム」

3. 2 知識ウェアハウスが提供すべき機能

(1) 情報の収集

インターネットに代表される社外のネットワークを自由にナビゲーションできる環境が整備されただけでなく、インターネット上を人に代わって代理人として作業を行うプログラム（エージェントという）が開発されている。この技術を用いる

ことで、インターネット上で情報を自動的に収集することが可能である。

高速な処理速度を持った情報処理により、インターネットに登録されているホームページの内容を全文検索することで収集・分類機能を持ったエージェントも実現できる。

(2) 情報の蓄積

前述の代理人プログラム（エージェント）が収集してきた情報を、そのままデータベースに蓄えるだけでなく、利用者の観点や視点に応じた情報の要約・抽出を行った後、情報を蓄積する。

また、インターネット上のあるページやイントラネット上のあるディレクトリを監視しているエージェントを実装することで、最新の情報がインターネットあるいはイントラネットに登録された時点で情報を収集する機能を実現できる。

(3) 情報の共有

高機能な DBMS を使うことでマルチメディア情報を取り扱い、収集・蓄積した情報をイントラネット上で公開したり、自然言語による検索や類似検索などの検索機能を充実することで、企業内での共有化を実現できる。

3. 3 知識ウェアハウスによる成果

(1) 意思決定支援

開発業務におけるアシスタント（あるいはコンサルタント）に匹敵する程度の知識を持ち、技術者の問い合わせに対する迅速かつ確かな回答を提示することで意思決定を支援する。

(2) 生産性の向上

適用技術・手法や開発上の問題点・解決方法などを技術情報としてデータベース化した知識ウェアハウスを開発し、全部門から参照可能とする。これにより、計画立案・運営・使用技術のノウハウを参考にして、無駄な調査等をなくしたり発生する可能のある傷害等を未然に防ぐことで生産性の向上を図る。

(3) 新たな知識の創造

「新たな知識の創造」とは、

・大量のテキスト情報の中に潜む共通の規則や特徴を抽出すること

- ・あるいは逆に特異な情報を抽出すること
- ・代替案を提示すること

であると定義する。具体的な例を示すと、大量の新製品情報の中で共通する特徴・特異な機能を抽出したり、あるいは技術者がある問題の解決方法として適用を試みる技術に対して、代替案を提示する。

3. 4 知識ウェアハウスとデータウェアハウスの類似点・相違点

(1) データウェアハウスとの類似点

提唱者ビルインモン氏の定義によると「目的別に編成され、統合化され、更新されない、時系列であるデータの集合であり、経営上の意志決定を目的とするデータベース」である。

(2) データウェアハウスとの相違点

まず、データウェアハウスが基幹系データベースを対象とし、その中に登録されている数値情報や記号情報などを取り扱うのに対し、知識ウェアハウスは自然言語で書かれたテキストを対象とする。次に、データウェアハウスが意志決定の迅速化・的確化に重点を置くのに対し、知識ウェアハウスは複数の情報源から得られた情報を統合することで新たな知識を創造することも目標としている。

4. 知識ウェアハウス実装のための基盤技術

知識ウェアハウスの構成は図1のようになる。

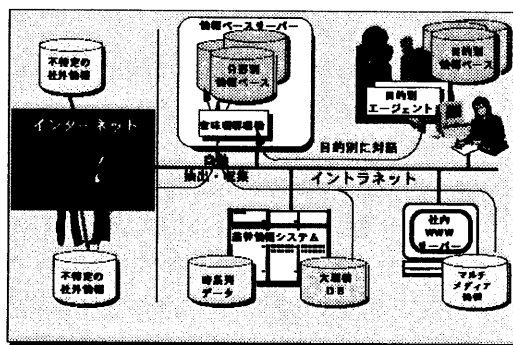


図1. 知識ウェアハウスの構成

●インターネット上の不特定な社外情報（各社のホームページから参照できるカタログ情報など）

●イントラネット上の参照可能な大規模 DB（社員
の技術履歴など）

●知識として蓄えられた情報ベース

●社内外のデータを情報ベース化するエージェ
ントから構成される。知識ウェアハウスを実装す
るために必要な技術を以下に説明する。

（1）データベースエンジン

高速／高負荷耐性の OLTP サーバーおよび柔軟
なオブジェクト指向モデリング機能を持ったデ
ータベースエンジンとして、MATISSE を利用す
る。MATISSE は、以下の特徴を持つ。

●厳しい稼働環境耐性（フォールト・トレラント、
機密管理、24時間ノンストップ運転）

●高度なトランザクション・スループット・ハン
ドリング

●ハイパーテキストとハイパーメディアナビゲ
ーション

●意味ネットワークに基づいた柔軟なモデリン
グ

（2）イントラネット構築技術

多くの企業内では Web サーバーを中核とした
イントラネットが急速に構築されつつある。知識
ウェアハウスの Web インターフェイスを構築す
るツールとして、MATISSE-Web Generator を利用
する。MATISSE-Web Generator は、MATISSE をパ
ックエンドとする Web インターフェイス構築ツ
ールである。HTML を拡張した HQL (Hypermedia
Query Language) により MATISSE データベースと
のインターフェイスを持つ。

（3）分野別情報ベースと目的別情報ベース

従来の情報処理では、意味関係の記述が十分で
なかったため情報の理解、類推、帰納推論、仮説
推論などの高度な処理が困難であった [1]。こ
のような問題を解決するために情報の意味的関
係を自動的に構造化する自己組織化の理論が考
案された [2]。この理論を適用し、動的な概念
構造を持つデータベースを情報ベースと呼ぶ。

知識ウェアハウスには、分野別情報ベースと目
的別情報ベースの2つの情報ベースがある。前者
はある目的の下で収集された情報ベースであり、

後者は分野別情報ベースを参照する検索者の視
点や観点を表現したものであり、個人用のシソー
ラスが相当する。

概念構造を自動的に作成する取組みとして、テ
キスト情報から同義関係と階層関係抽出、階層関
係および関連関係抽出、意味関係抽出などのアル
ゴリズムがあり、目的別情報ベースを自己組織的
に生成するのに適用される。

（4）ドキュメントマイニング

ドキュメントマイニングとはデータマイニン
グの関連技術であり、データマイニングが数値あ
るいは記号化された情報を対象とするのに対し、
ドキュメントマイニングは大量のテキスト情報
からその中に潜む知識を発見し、自動分類や要約
に利用しようとするものである。ドキュメントマ
イニングの要素技術を以下に説明する。

●類似検索技術

任意の日本語／英語テキスト全文情報に対し
て、構造化された辞書（シソーラス）を用いて、
用語の出現頻度で重みづけられた類似度計算を
行い、類似性の高い文章を抽出する技術である。
類似度計算には、事例ベース推論でよく用いられ
る最近傍アルゴリズムを用いている。シソーラス
の作成を支援する仕組みは、構造を持たない用語
リストをテキスト情報から自動的に作成すること
で実現している。3)の自己組織化アルゴリズム
を適用することでシソーラスの自動生成が可能
である。

●自動要約技術

特定専門分野の情報に対して、英語論文テキ
ストから技術情報の抽出・要約・構造化・統合を自
動的に行なう技術である。ここで中心となるのは、
技術情報の抽出法と構造化法を一体化したドメ
イン知識のパッケージである KP (Knowledge
Pieace) であり、KP の組み合わせによる自動要
約が可能である [3]。知識処理技術の一つであ
るオントロジーを導入することにより、他の専門
分野への拡張や精度向上が可能であると考えら
れる。

（5）高度な意味処理を行う知的エージェント

知識ウェアハウスの開発に必要なエージェントは、ネットワーク内に存在する（自立的な）ソフトウェアの総称とし、「ソフトウェアエージェント」あるいは「ネットワークエージェント」と呼ばれるエージェントのことである。

知識ウェアハウスで必要とするエージェントは、文字どおり代理人として検索要求者に成り代わって作業を行ってくれる知的なエージェントという意味で、「知的エージェント」とした。知的エージェントが、知識ウェアハウスで果たす役割は、

- 情報収集エージェント（ネットワークからの情報ソースの自動収集）

- 情報抽出・分類エージェント（収集された情報ソースから情報の自動抽出・自動分類）

- ニュースエージェント（電子ニュースの自動分類・自動発送）

である。

5. 知識ウェアハウスの実装例

本章では、情報処理に関する技術情報を対象とする知識ウェアハウスのプロトタイプシステムを説明した後、開発予定の実用システムについて述べる。

5. 1 対象とする技術情報

(1) 個別のメーリングリスト

多数の技術情報が企業内の特定のメーリングリストで転送されている。

- メーリングリストのメンバーが必要とする技術情報と転送される情報が不一致である。

- サブジェクトのみで判断した結果、参考にする必要がないと判断される。

のいずれかの理由で目を通さずに消去されることがある。

(2) BBS での情報交換

BBS 上で技術情報について問い合わせると、電子掲示板で討論されるか、直接質問者に回答が返送される。しかし、一般に同様の質問が繰り返し発信されることが多いため、FAQ (Frequency Asked Question) と呼ばれる解答集が既にまとめ

られていることもある。その場合、質問者は FAQ に含まれるかどうかを前もって検索することが望まれるが、どのような FAQ がどこにあるのかという基本的な検索を支援するツールがない。

(3) Web で発信されている情報

Web 上では、ソフトウェアやハードウェアの新製品の情報が発信されているが、検索エンジンの精度が高くないため、必要な情報を見つけるのに労力を要したり、あるいは抽出されるべき情報が検索対象として抽出されないこともある。

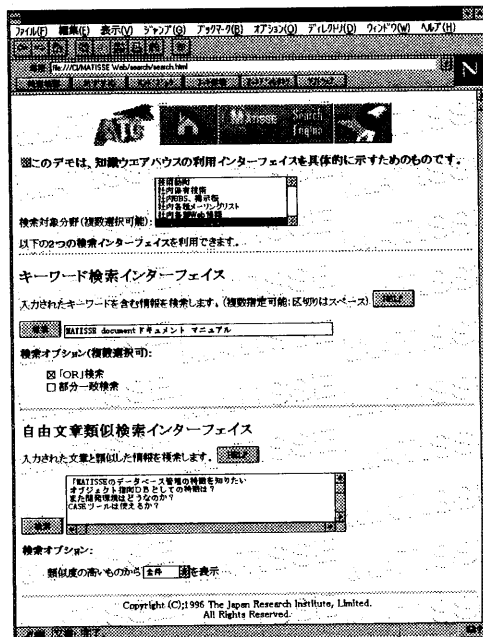


図2 検索条件指定

5. 2 知識ウェアハウスプロトタイプシステム

現在の保有技術で開発したシステムを説明する。

サーバー側は、MATISSE-Web Generator を用いバックエンドデータベースは MATISSE を利用している。また、クライアントに必要なソフトウェアは一般的な Web ブラウザのみである。

図2は検索対象分野と検索方法の指定を行う Web ページである。検索対象分野は、知識ウェアハウスにおける分野別情報ベースに対応しており、この例では【MATISSE 利用情報】を指定し

ている。また、検索方法はキーワード検索と自由文章に基づく類似検索の両方を指定できる。キーワード検索インターフェイスは、MATISSE が標準で提供しているフルテキストサーチを用いるものであり、高速な部分一致検索が可能である。自由文章類似検索インターフェイスは、入力された文章から対象分野別のシソーラスを用いてキーワードを自動抽出し、これを用いて類似検索を行う。

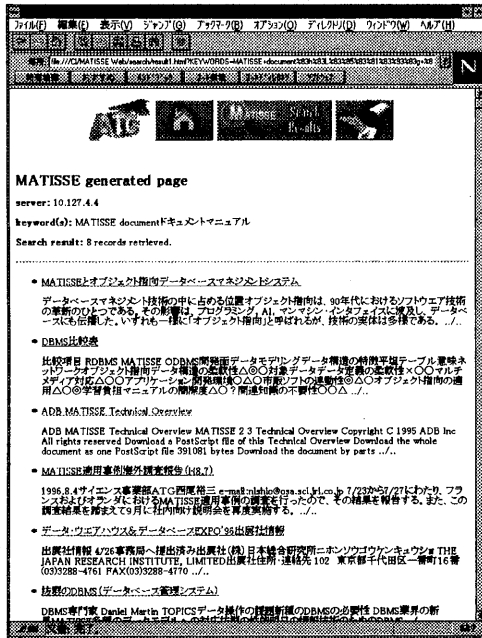


図3 検索結果 (キーワード検索)

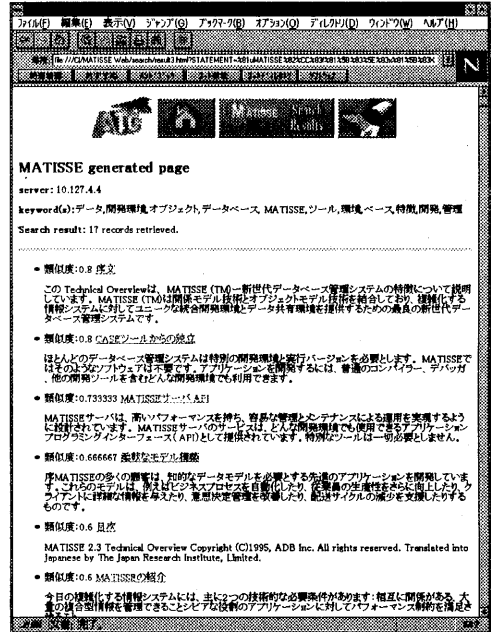


図4 検索結果 (自由文章類似検索)

図3がキーワード検索インターフェイスの検索結果であり、図4が自由文章類似検索インターフェイスの検索結果である。どちらの検索結果もオリジナルの情報源のURLとリンクされており、元の情報を参照できる。キーワード検索を準備していることにより、従来と同様のインターフェイスと、自由文章類似検索によるさらに高度なインターフェイスの両方を提供できる。

このような利用環境は、従来のような分散したWeb個別の検索エンジンを用いて実現すると、非常に処理時間がかかり実用的でないものとなることが容易に推測される。

5.3 開発予定の実用システム

このプロトタイプシステムでは、技術情報の内容が分野別に整理・収集されており対応する各分野別シソーラスと共にデータベース上に登録されている。4章で述べた自己組織化技術、知的エージェント技術、ドキュメントマイニング技術を適用することにより、以下のような実用システムが開発できると考えている。

- (1) 最新情報の自動ダイレクトメール (ニュー

スエージェント)

ユーザーが指定した分野の最新情報を自動的にダイレクトメールする。指定する内容は、例えば定期的に調査したい技術動向などの複数のキーワードや新聞記事のような自由文章でもよい。メールで送られる内容は、オリジナルの情報の URL や、内容の要約などである。全ユーザーの指定内容をニュースエージェントサーバーに保管することにより、一度受け取った内容は再送されないようにできる。

上記のような利用環境により、興味のある情報を漏れなく収集し再利用することが可能になり、しかも情報収集にかかる労力はほとんど不要となる。

(2) 最新の Q&A 情報を使った PC ヘルプデスク

知識ウェアハウスによる PC ヘルプデスクは、ユーザーが質問した時点における最新の情報を用いて解答を行うことが可能になる。ユーザーは自然文章で質問し、ヘルプデスクエージェントはユーザーの質問に関連した情報の中から、最も似た質問への解答を見つけ出す。質問/解答の情報は、各種の BBS や Web の掲示板などから自動収集することにより、常に最新の Q&A 情報が準備されていることを保証できる。十分に似た内容の質問が見つからない場合に備え、この質問を適切な情報収集場所（関連情報が最も議論されているネットワークなど）へ自動的に掲示する機能も必要である。

6. まとめ

知識ウェアハウスを提案し、情報処理に関する技術情報を対象としたプロトタイプシステムを構築した。現在のプロトタイプシステムは、知識ウェアハウスが提供すべき機能の一部を実装したものであり、今後大学との共同研究をすすめ、機能を充実する一方、社内外にシステムを公開し、評価を行う予定である。

知識ウェアハウスを用いて情報利用の高度化を行うためには、情報のソースがイントラネット

上で参照できる形式で保存され、情報提供者の合意の元で知識ウェアハウスに登録・利用できることが前提条件となる。情報公開/利用に対する知的所有権のような法的な検討が必要であり、情報提供者に対する制度上の優遇措置についても検討をしなければならない。

参考文献

- [1] 尾関, 藤原 et al, "情報学の概要とその現状", 情報処理, Vol.36, No 10, Oct. 1995
- [2] Y.Fujiwara, "The Model for Self-Structure Semantic Relationships of Information and Its Advanced Utilization", International Forum Information and Documentation, vol.19, No.2, April 1994
- [3] Toshiyuki MATSUO and Toyooki Nishida: Intelligent support for construction and exploration of advanced technological information space. Proc. 4th Int.Conf.on the Int. Society of Knowledge Organization pp.336-345(1996)

知識ウェアハウス™ および Knowledge Warehouse™ は、株式会社日本総合研究所の登録商標です。