

分散環境における株関連情報システム

吉田 武俊 有馬 淳 吉野 利明 寺本 良明

(株)富士通研究所
コンピュータシステム研究所

我々は、ネットワーク上に分散する異種情報源を統合し、ユーザの平易な要求や仕様に基づいて情報源を選択し、抽出した内容を関連付け表示する事により知識の創造支援を行うエージェントシステムの研究を行なっている。本稿では、その最初の段階としてエージェントによる情報統合部分の実装に焦点をあて、株に関する情報の関連付けを例題としその試作を行なったので報告する。本システムは、Webから動的に株価情報を取りだしチャート表示し、新聞記事等と関連づけてユーザーに表示する。今後、本例を一般化する事により、応用域を広げ、より自由な表現と関連づけを平易に行なうための記述方法を探る予定である。

Related Information System on Stock Market Under Distributed Environment

Taketoshi Yoshida Jun Arima Toshiaki Yoshino Yoshiaki Teramoto

Computer Systems Laboratories, Fujitsu Laboratories LTD.

We study a software agent system which integrates heterogeneous information sources distributed on the network, selects information sources among them based on a simple specification given by a user, then relates the extracted content, and displays in a coherent form. In this paper, we present our experimental system focused on the methodology of integrating information about stock market. This system dynamically extracts the price of a required stock, represents it in a graph and represents relevant information to it (ex. news, outline...). In a future work, we will expand applied region by generalizing this example and study a general method to indicate which contents are related and how to display them simply.

1. はじめに

近年、多種多様な情報ソース(RDBやWEBなどの(半)構造化データ)が共通の情報流通基盤(インターネットなどのネットワークインフラ)を用い、特定のネットワークドメイン内外で共用されるようになってきた。情報のユーザにとって、利用可能な情報ソースが容易に入手できる事は、情報の分析範囲や深さを広げ、分析の結果からもたらされる判断の適切さを向上させられる期待がある。これらの目的の為に、多種多様なデータ構造で管理されている情報ソースを統合管理するシステム内で共通化し、検索結果に含まれる一部の必要な情報を可能な限り一般的な要素として機械的に取り扱えるようにする事で情報の利便性を向上させる研究が盛んに行われている。

この様に、分散した多種多様な情報ソースを統合する枠組みとして、各情報ソースを管理するエージェントを設定し、検索要求の処理をしたり、これらのエージェントを一元的に管理するエージェントを設定する事により、分散された情報ソースを一つの情報ソースとして取り扱えるようにしたりする枠組みが研究されている([1][2][3][4])。本研究の目的は、この様な情報ソース統合の枠組みを用いて、ユーザの平易な要求や仕様に基いて情報ソースの選択、関連付け、表示を行う方法を探ることである。

本稿の構成は次の様になっている。本研究で採用しているエージェントを用いた情報ソースの統合システムの枠組みを紹介し、本研究への適用について述べる。次に、情報の要求から関連付けを行う方法について述べ、その問題点をあげ検討を行う。

2. 統合の枠組み

本研究で採用している情報ソースの統合研究として、ファシリテータを用いた情報統合の研究がある[1][2]。そこでは、情報ソースをシステムへ統合させる為のデータベースエージェント、各デ

ータベースエージェントを統合し一元的に管理するエージェントであるファシリテータ、ユーザとファシリテータ間のインタフェース機能を提供するユーザエージェントから構成されている。

(1) データベースエージェント

データベースエージェントが管理対象とする情報ソース(主にRDB等)固有のインタフェースをファシリテータがアクセス可能なインタフェースに変換し、また情報ソースのプロパティ情報をファシリテータに**アドバタイズ(宣伝)**する。例えば、ある情報ソースが会社概要に関する情報を登録していたとする。その情報が、会社の“社長名”、“代表電話番号”、“住所”、“役員”などの項目から構成されていた場合、統合システム内におけるその情報ソースの識別子と共に、これらの情報をファシリテータにアドバタイズする。アドバタイズされた情報は、ファシリテータがユーザエージェントからの(検索)要求に適合する情報ソースを選択する判断基準として用いられる。

(2) ファシリテータ

ユーザエージェントから送られてくる(検索)要求を解析し、データベースエージェントからアドバタイズされた情報を元にその要求に適合する情報を管理するデータベースエージェントが解釈可能な要求形式に変換し送信する。各データベースエージェントからの結果は、項目ごと(例えば、会社の“社長名”や“住所”など)にデータをマージしユーザエージェントへ送信する。

(3) ユーザエージェント

ユーザからの(検索)要求を内部インタフェース形式に変換し、ファシリテータへ送信する。また、ファシリテータから送信されてくる(検索)要求に対する結果を、ユーザの表示ツールのデータ形式(主にHTML)に変換し表示させる。

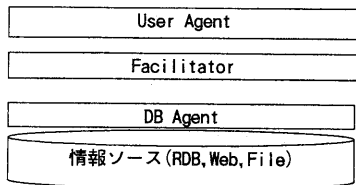


図1：情報統合システムの構成

(4) 適用

本研究では、情報の統合の枠組みにこれら3つのエージェントからなる枠組みを簡易化し適用する方法を用いている。主に株に関する情報ソースとして、採用した情報ソースとその媒体の関係を表1に示した。

表1：情報ソースと媒体

		情報媒体		
		Web	File	RDB
情報 ソ ース	株価	○		
	新聞(読売, 産経)			○
	会社概要		○	
	地図		○	

ユーザの要求を該当するデータベースエージェントに送信するファシリテータに、情報の関連付け機能をもせた構成を採用している。

ユーザエージェントは、平易な要求仕様に応じてHTMLで記述した要求入力フォーマットをCGI経由にてユーザのブラウザに表示し、要求事項を満たされたクエリをやはりCGI経由で受け取り、内部データ軽視に変換後、ファシリテータへ送信する。

3. 情報の関連付け

新たな知識の創造を支援する事を目標に、独立した情報ソースの情報の要素化を行い、情報ソースもしくは情報の要素の関連付けに関する一般的な方法論を探っている。そのアプローチとして株に関連する情報に限定し、実験システムを作成した。ここでは株関連情報システムについての機能

のうち、情報ソースの情報の要素化、ユーザの要求表現、情報の関連付けの各方法、動作概要、機能構成の概要を以下に述べる。

(1) 情報の要素化

情報の要素化とは、各情報ソースに対して次の3つの処理を行う事である。第一に、ユーザの要求に対して情報ソースから結果として示される情報を、更に小さな情報の要素として分解し、機械的に取り扱えるように“タグ付け処理”を行う事。第二に、要求部分に該当する情報だけを抽出する処理(以下、“部分抽出処理”)を行う事。第三に、情報ソースから情報を検索する為に情報ソースに与える情報(例えば、“会社名”とか“日付”などの情報であり、以下これらを“検索キー”という)を内部的表現に統一する(例えば、“会社名”, “銘柄名”, “企業名”を“会社名”に統一にする)処理(以下、“類義語整合処理”)を行う事である。

“タグ付け処理”は各情報ソースに対して行われている。例えば、会社概要情報ソースはラベル(会社名、役員名、代表電話番号等)とその値(F株, F太郎, 03-123-...等)の組み合わせでテキスト形式として登録されている。これを明示的にラベルとその値とに分解し要求元に結果として返す。

“部分抽出処理”は、株価情報ソースに行う。例えば、株価情報ソースはWebに公開されているHPから、必要な情報(検索キーとして“会社名”と“期間”が定義されている)を入力後取得されるものであって、株価情報が記述されている部分に対して部分抽出処理がなされる。株価情報は、HTMLを用いて表形式に表されており、バーナー広告など実際には必要の無い情報を除き、タグ(日付、始値、終値等)とその値(価格)の組み合わせに変換し要求の結果として戻される。

“類義語整合処理”は、全てのタグ情報と検索キーに対して行われる。例えば、情報ソースに“会社名”としてタグ情報の候補が現れているのに対

して、実際の情報ソースの検索キーには、“企業名”とされている。これらは内部データに表現するに当たって“会社名”と表現する事により、タグ情報と検索キーとのマッチングを行っている。現システムでは、これらの条件処理は固定的に記述している。

(2) ユーザの要求表現

ユーザの要求をシステムへ伝える為の表現は、簡単なクエリーフォームをユーザに示し、必要な情報を入力してもらう事により、ユーザの要求を表現させている。この時、関連させる情報ソースは、システムが自動的に選択するのでなく、アドバタイズされている情報ソースからユーザが選択する。選択範囲は、新聞情報ソースに関して、読売新聞、産経新聞からの選択、会社概要情報ソースについては関連させるか否かの選択である。地図情報ソースの選択に付いては会社概要情報ソースが選択された場合、自動的に選択される。

(3) 関連付け

各情報ソースの関連付けは、次の2つの基本的条件を満たすように行われている。

- (1) 情報ソースを検索した結果である情報に表れるタグが、別の情報ソースの検索キーとマッチする事
- (2) ユーザが入力する情報（例えば会社名や日付）に付けられているタグが情報ソースの検索キーとマッチする事

これらの条件にマッチした情報要素にはリンク情報を生め込む事により情報ソース間の物理的な関連付けが行われる。

(4) 動作概要

図2は、ユーザの要求情報に基づいて関連付けされた情報の例を表した図である。この例では、

ユーザの要求として次の情報が与えられる。

- ・社名：“富士通”、
- ・期間：“1997年の10月1日から11月1日”
- ・関連付け新聞：読売新聞、産経新聞

この内、社名及び期間は株価情報ソースの検索キーとして用いられている共に新聞情報ソースとして用いられる。この時、新聞情報ソースの検索結果において、“富士通”に該当する記事があった場合チャートの日付部分が記事の無い日付とは異なる色で表されると共に、その記事へのリンク情報

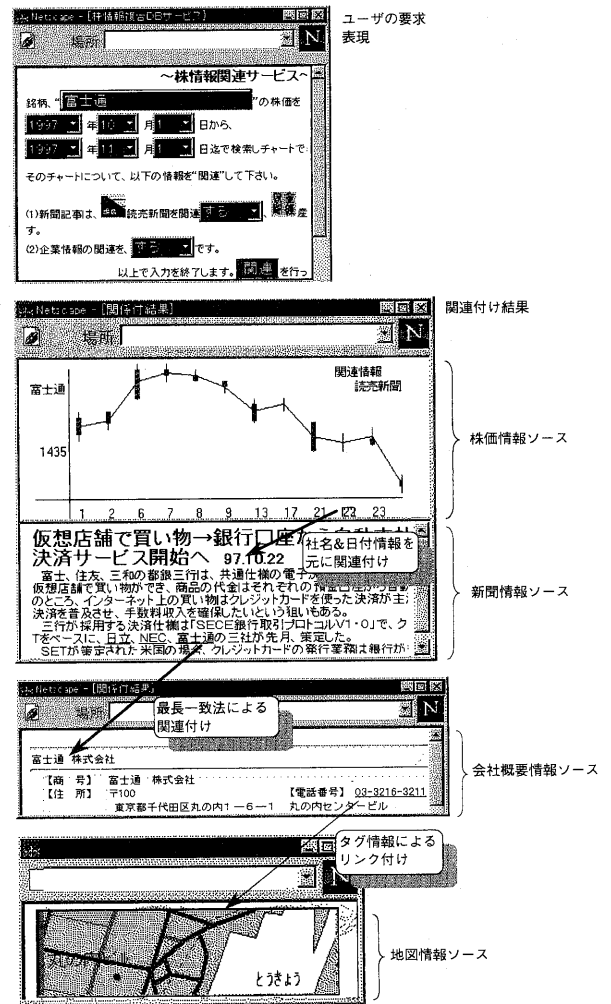


図2：関連付けされた情報の例

報が埋め込まれる(例えば図2の"4月10日")。この部分をクリックすると、その記事が表示される。

ここで、新聞記事に"富士通"の文字が現れると、会社概要情報ソースに登録されている社名と最長一致法を用いてマッチングを行い、該当する会社概要情報へのリンクが埋め込まれる。

会社概要情報が表示される時には、タグ情報である"代表番号"からその値である"03-123-XXXX"を取り出し、地図情報ソースの検索キーとし検索を行い該当データへのリンク情報を埋め込む。会社概要情報は、その情報を構成する項目をタグ情報としてシステム内部で捉えており、タグ情報が検索キーとなり得る情報ソースへの検索に用いられる。

このように複数の情報ソースを適宜に関連付ける事により、本来独立した情報ソースだけでは見えなかった知識、即ち株価変動要因と言う知識創造をユーザに支援させる新たな情報を生成する事が可能となったと思われる。最も実用上、より価値のある変動要因に関する知識創造を支援するには、更に多くの適切な情報ソースを関連付ける必要がある。

(5) 機能構成

図3は、本システムの機能構成を示した図である。

- ・ ファシリテータ

ユーザエージェントから送られてくる要求を分析し、データベースエージェントへの検索キーへ分析した値を割当てて送信する。データベースエージェントから送られてくる結果を、関連付けの基本条件に沿って纏め結果をユーザエージェントへ送信する。

- ・ データベースエージェント

Web上で公開されているHPなどの構造的なデータでない情報ソース(株価情報ソースや地図情報ソース)と、構造化された情報ソース(会社概要情報ソースや新聞情

報ソース)に対応づけてデータベースエージェントを設けた。両エージェントとも情報の要素化機能が実装されている。

- ・ ユーザエージェント

要求の表現を提示し必要な情報の獲得及び、関連付けされた結果をブラウザに表示可能な形式(HTMLなど)に変換する。

- ・ 類義語整合処理辞書

各情報ソースで情報の要素化を行う対象となるタグ部分の表現とシステム内部の表現との対応が記述されている。例えば、情報ソースでの"企業名"は内部では"会社名"である事などである。

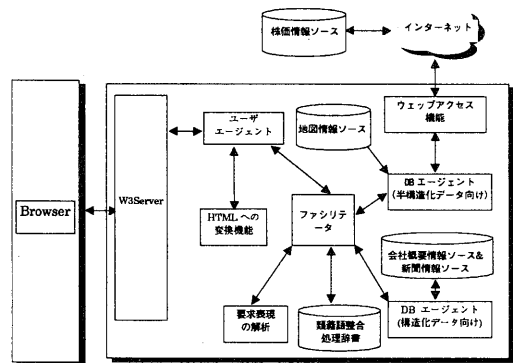


図3：機能構成

4. 課題

現システム作成を通じ、次の3つの技術課題が抽出された。

- (1) タグ情報にマッチした複数の情報ソースの中から、関連付けをする為の適切な情報ソースを選択する方法の工夫。

例えば、ユーザは株価の変動要因を探る要求を出し、株価のグラフが表示されており、このグラフに関連付けする情報ソースが、株価のタグ情報より新聞情報ソースと会社概要情報ソース

が候補としてあげられたケースを考える。ユーザは変動要因を要求している為、会社の概要に関する情報をリンクしても変動要因は探れない。それより新聞記事をリンクすればどのような理由で変動が起こったのかを探る事が可能となる。この様に、ユーザの要求からリンク元の情報の特徴という文脈を考慮した情報ソースの選択が必要と思われる。

- (2) ユーザの一般化された要求及び仕様に関する記述の工夫。ユーザが関連付けされた情報をどのような書式で表示して欲しいのかを記述する方法である。例えば、株価の表示としては、表形式で表示する事も考えられし、グラフで表示する事も考えられる。このような要求を一般的な記述方法を用いて表す必要性が有ると思われる。
- (3) 情報要素の抽出処理を宣言的な記述をデータベースエージェントに与える事で情報ソースのシステムへの統合を容易にする工夫。現在、各情報ソース毎にデータベースエージェントを作成している。しかし、機能的に共通する部分が多い為、情報ソース固有の処理部分をファイル等に記述するしデータベースエージェントに取込む事で固有のデータベースエージェントを実現する必要性が有ると思われる。

情報ソースの情報管理の仕方として必ずしも検索キーが割当てられているケースだけではなく、単にディレクトリにファイルが置かれそのインデックスすらないケースも多々見られるようになってきた。このようなケースに現手法の対応可能性を検討することも今後の検討課題である。

5. まとめ

本稿では、事例に基づいてユーザへの要求に基づいてネットワーク上に分散する情報ソースを選択し、抽出した内容を関連付け表示する株関連情報システムについて述べた。情報の関連付けとは、情報の要素をユーザの平易な要求や仕様に基づいて、独立した情報ソースだけでは選られない新たな知識を創造する支援をする事であり、情報の要素とそれを纏め上げる概念体系の組み合わせが必要である。我々の事例では、情報の要素化とそのタグ情報に基づいた関連先の情報ソースの選択方法により株価の変動要因究明という知識創造支援を示せたと思われる。

今後の課題としては、概念体系の整備を含めこれらの手法を一般的なものに広げる事により、多様な知識創造支援に対応して行く事があげられる。

参考文献

- [1] Michael R. Genesereth, Steven P. Ketchpel: "Software Agents", Communications of the ACM, July, pp.48-53, 1994
- [2] 丸山文宏, 益岡竜介, 菅坂玉美, 佐藤陽, 大竹聡, 渡辺光好: "SAGE(Smart Agent Environment)-仮想カタログ-", 第54回情報処理全大, vol.3, pp.129-130, 1997
- [3] H. Garcia-Molina, Y. Papakonstantinou, D. Quass, A. Rajaraman, Y. Sagiv, J. Ullman, V. Vassalos, J. Widom: "The TSIMMIS approach to mediation: Data models and Languages". In Journal of Intelligent Information Systems, 1997.
- [4] Bressan, S. and Bonnet: "Extraction and Integration of Data from Semi-structured Documents into Business Applications", Conference on the Industrial Applications of Prolog, 1997