

ネットワークからの知的情報収集・統合

Information Gathering and Integration from Network Information Sources

武田英明

Hideaki Takeda

奈良先端科学技術大学院大学 情報科学研究科

Graduate School of Information Science,

Nara Institute of Science and Technology

1 はじめに

今日インターネットの利用の拡大によって、容易に入手可能な電子化された情報は極めて莫大になっている。例えば、検索エンジンである Alta Vista¹では 28 万サーバの 3000 万の WWW ページを収集しているという。これはネットワーク化以前からの文書作成など個人 / 組織レベルでの情報電子化の経験の蓄積が、WWW に代表される簡便な情報公開手段の普及とあいまって実現したものと思われる。

現在の状況は、これまでの電子化の流れとは大きく異なっている。それは、これまでの電子化された情報は目的指向であったためその利用法は自明であった。すなわち、電子化情報源をつくるということは、なんらかの目的があり、その目的のために情報源の設計がなされていた。この場合、情報利用者が各情報源の目的を理解するのは容易であり、情報提供者も情報利用者も相互に目的を共有しているということを前提としてた。このための情報源の利用において、その目標と方法は自明であり、したがって情報源の利用すなわち情報検索であった(図 1 参照)。利用者のすべきことは情報源を直接アクセスするか、あるいはアクセスの仕方をまとめたスキーマのようなものを作成して利用する程度であった。

WWW では多様かつ広範な領域の情報が存在している。ここで問題となるのは単に情報の多様性でなく、情報源の目的が自明でない点である。WWW では、これまでの情報化の流れに沿って行ってきた我々の業務や生活の情報化した側面を情報源として利用することを可能にした。このことは、必ずしも情報源がその目的をもっているか自明ではないということの意味してい

¹<http://www.altavista.digital.com/>

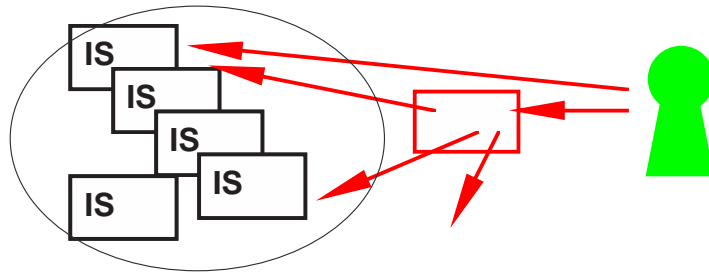


図 1: これまでの情報源の利用方法

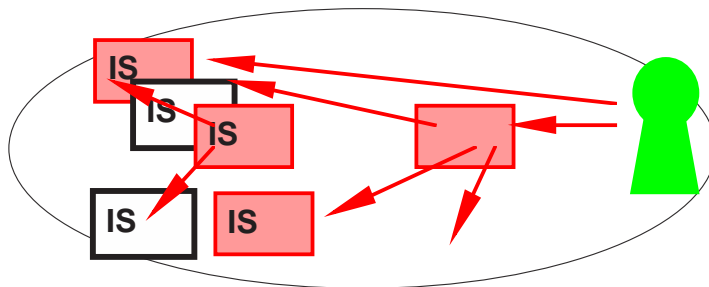


図 2: ネットワーク情報源の利用方法

る。ある場合は暗黙的に目的が存在するし、またあるときはそもそも目的が存在しない場合もある。このような場合、当然情報利用者はこれまでのように情報源の目的を理解すれば情報が得られるというわけではなくなる。このような受動的な立場ではなく、自らが情報を統合化する作業をするという能動的立場が必要になる。この意味で情報源の利用とは、単に情報検索だけではなく、情報収集、分類、抽出など含めた情報統合となる（図 2 参照）。本稿では、上記の意味での情報源の利用をネットワークからの知的情報統合として位置付ける。

2 問題設定

問題領域としては、WWW に代表されるような多様性（形式面、内容面）、分散性、大規模性を扱う情報源群の利用とする。形式面での多様性とは、データベースといったフォーマットが厳密に定義されるものから、EメールやHTMLのように内容的には自然言語であるものまで、電子化される方法が多様であるということである。内容面での多様性とはWWWのように製品情報から個人の日記までといった多様性を指す。分散性とは、情報のある場所が分散しているだけでなく、情報の責任者が分散しているという点も重要である。大規模性はここではそれぞれの情報源が大規模であるということよりも、情報源の数が多量であることが重要である。

本稿では上記のような性質をもつ情報の利用に関して、ネットワークを用いている研究を中心に概観するが、必ずしも現状でWWWやネットワークを用いないものも紹介する。これは、上

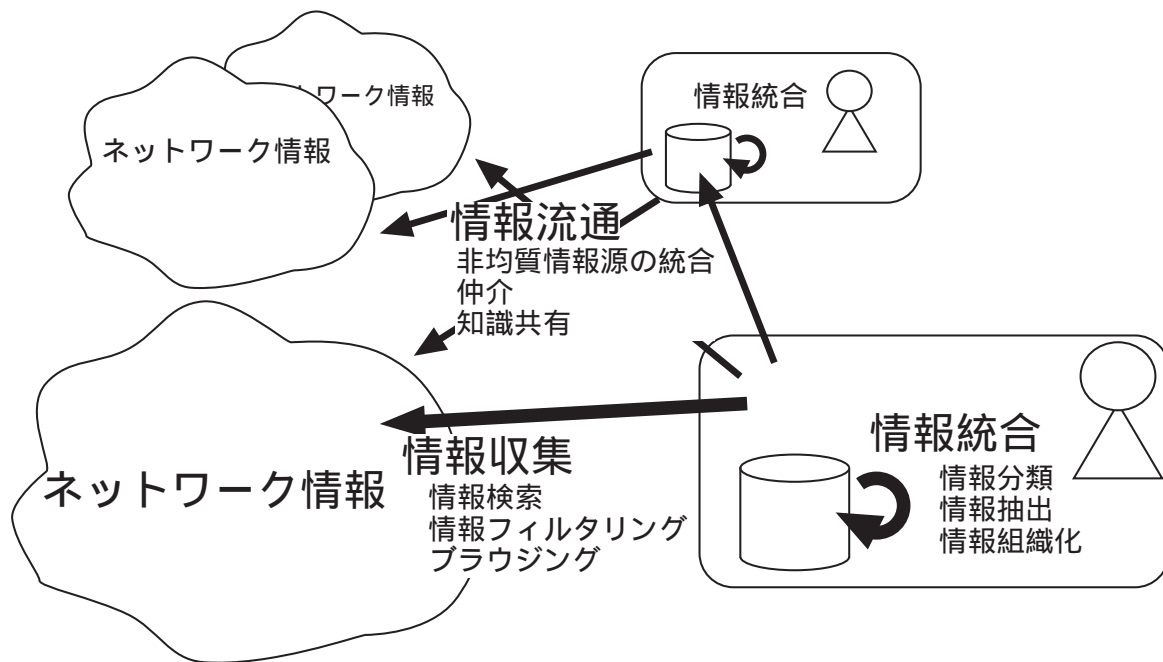


図 3: ネットワークを利用した知的情報統合の枠組み

記のような性質をもつ情報を扱うのであれば、ネットワークを利用するか否かということは、技術的には本質的ではないと考えたためである。

3 対象分野

ネットワークを用いた知的情報統合には、これまで独立した分野として研究されてきた様々な分野が関係してくる。そこで、図 3はネットワークを用いた知的情報統合に関わる概念を模式的に示したものである。

まず、大きな 3 つの分野としては、情報収集 (Information Gathering)、情報統合 (Information Integration)、情報流通 (Configuration of Information Sources) があげられる。これらは情報源 (集合) に対する処理主体の立場によって規定される。

まず、ネットワークからの情報収集 (information gathering) としては、情報検索 (information retrieval)、情報フィルタリング (information filtering)[50]、ブラウジング (browsing) の方法が挙げられる。情報収集では、主体は情報源とは情報を取りにいく関係のみ規定される。ここでは主体が情報源にどうアクセスするかが主な課題となる (図 4)。複数の情報源のある場合の情報源間の関係は自明ではない。

情報の統合としては、情報分類 (information classification)、情報抽出 (information extraction)、情報組織化 (information organizing) などが挙げられる。情報統合では主体と中心として、個々の情報源もつ情報をどう結び付けるかが課題である (図 5)。この場合、複数情報源間の関係が当然考慮されるが、むしろ情報源間の関係より、それらが持つ情報間の関係が重要であ

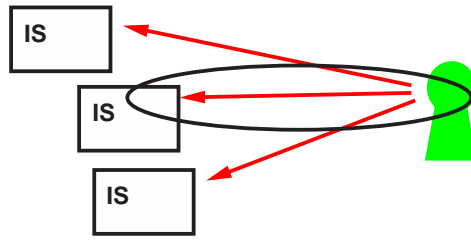


図 4: 情報収集のモデル

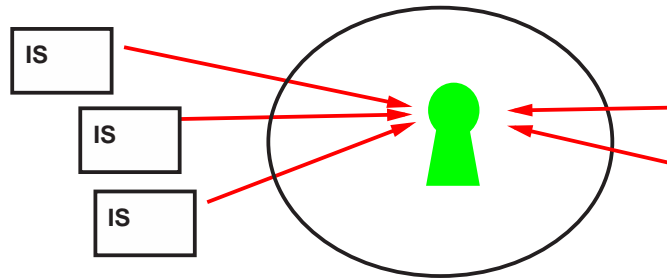


図 5: 情報統合のモデル

る。

情報流通においては、情報源や情報利用者が多数かつ異なる場合、どのようにして情報源と情報利用者の関係を適切につくることができるかという問題である。具体的には情報源の表現、情報収集のプランニング、仲介などがテーマとなる。情報流通では、情報源の集合における情報源間の関係を明らかにすることが課題である(図6)。この場合、主体は情報源の集合の外にある。

4 ネットワークを用いた知的情報統合を可能とする技術

前章で述べたようにネットワークを用いた知的情報統合には様々な研究分野が関連してくる。そのなかで、特に鍵となる技術・研究は以下の点である。

柔軟なテキスト検索技術 特定の目的のために作成されてきたデータベースと異なり、ネットワークに存在する情報は、情報の品質と情報の記述形式が多種多様である。このような場合でも検索が可能な柔軟な検索技術が必要である。

頑強な自然言語処理 膨大かつ多様な情報の中からユーザの欲する情報を得るためにはキーワード検索だけでは不十分であり、一定の内容を理解する能力が必要となってくる。テキストを理解するという意味では自然言語処理であるが、ここで必要なのは詳細な内容理解ではなく、どういった内容が書いていあるかといった概略の理解である。通常の間文解析を用いているものもあるが(例えば[4])、対象が非均質であるがために汎用性が問題になる。こ

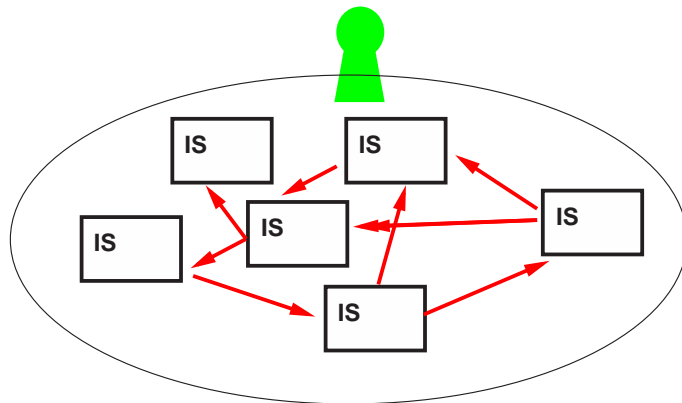


図 6: 情報流通のモデル

のためより浅いが頑強な自然言語処理、たとえば統計的な処理やテンプレートを併用する方法 (FAQ Finder[9][22][23])、共起関係を利用した方法 (例えば METIS[41][42]) などがある。

知識体系 広範な情報の関連性を知るためには単に情報源の情報を利用するだけでは難しく、背景知識が必要となる。このためには例えば語彙体系や概念体系の利用が考えられる。検索用概念体系としては例えば YAHOO²のカテゴリのようなものがあるが、既存の情報の体系化に焦点が当てられていて、かならずしもユーザのための概念体系ではない。汎用な概念体系を目指したものとしては IICA[29] のオントロジーがある。また、語彙体系としては WordNet[45] を利用しているものもある (FAQ Finder)。

ユーザ・モデル 予め用意されているようにしか動かないシステムは新しい情報やユーザの新たな利用法に対処できない。ユーザの利用に沿った情報収集や情報統合ができるようシステムが必要である。この一つの方法として一定の自律性をもちユーザの代行を行なうインタフェース・エージェントという考え方がある [46]。Maes の Learning Interface Agents[39][40] では、電子メールやスケジュールなどの作業において、観察による学習、ユーザフィードバックによる学習、教示による学習といった方法によってユーザの作業を学習させて、適切な作業支援できると報告している。

5 情報収集

先に述べたように情報収集の立場とは、ユーザの立場からどう情報源にアクセスするかという立場であり、情報検索、情報フィルタリング、ブラウジングなどの方法があげられる情報検索はユーザが欲する情報の仕様が明確でありかつ情報源が現にアクセス可能な時に有効である。情報

²<http://www.yahoo.com/>

フィルタリングでは、情報源が動的に変わる場合など現時点ではアクセスできない場合に有効である。ブラウジングは欲する情報が明確でない時有効な方法である。本稿ではこのなかでブラウジングに注目する。

5.1 ブラウジング

ブラウジングは検索と異なり、漠然とした目標しかもっていないときでも、情報収集が可能である。しかし、ブラウジングでは人間が常に選択を行なうため、人間の癖(「上から順に」、「寄り道」など)や人間の限界(逐次実行性など)によって、必要な情報にたどり着くことが難しいことがある。そこで、ユーザの主導権を維持しつつ、いかにユーザの行為を支援できるかが必要である。すなわち、ブラウジングにおいては、いかにユーザの好みを獲得するかが課題となる。この方向の研究として以下の研究を取り上げる。他にも Balabanović ら [3] の研究もある。

5.1.1 Web Watcher

Web Watcher[2] はカーネギーメロン大学の Armstrong らによって開発されたユーザのブラウジングからユーザの好みを学習するシステムである。Web Watcher の主な機能としては次のものがある。

- ユーザの宣言した興味に対して有用だと判断した現ページ中のハイパーリンクの推薦。
- ユーザの興味に従ってハイパーリンクの付加。

システムの構成としては、Web Watcher は Web サーバとブラウザの間に位置している。Web Watcher はユーザがブラウザを通じて Web サーバに要求したページを必要な改変をしてからブラウザに渡す。このため、ブラウザには依存しない。

図 7 で、冒頭に列挙された 10 のハイパーリンクは Web Watcher によって提案されたものでオリジナルのページはないものである。また、下の方の二つの四角印で囲まれたハイパーリンクは Web Watcher がこのページのリンクの中で次にたどるにもっともふさわしいと提案しているものである。

学習のアルゴリズムは以下の通りである。学習すべきものは、

$$LinkUtility : Page \times Goal \times User \times Link \rightarrow [0, 1]$$

である。すなわち、あるページにおいて、あるゴールのために、あるユーザが、そのページの中のあるリンクをどの程度評価するかということである。このうちユーザを除く部分をそれぞれ単語の集合で表現する。ページはハイパーリンクになっている部分のみを抽出の対象として、200 単語からの特徴ベクトル [51] で表現する。リンクはリンクになっている部分と対象として 200 単語、リンクを包含する章節のタイトルを対象として 100 単語の特徴ベクトルとする。ゴール

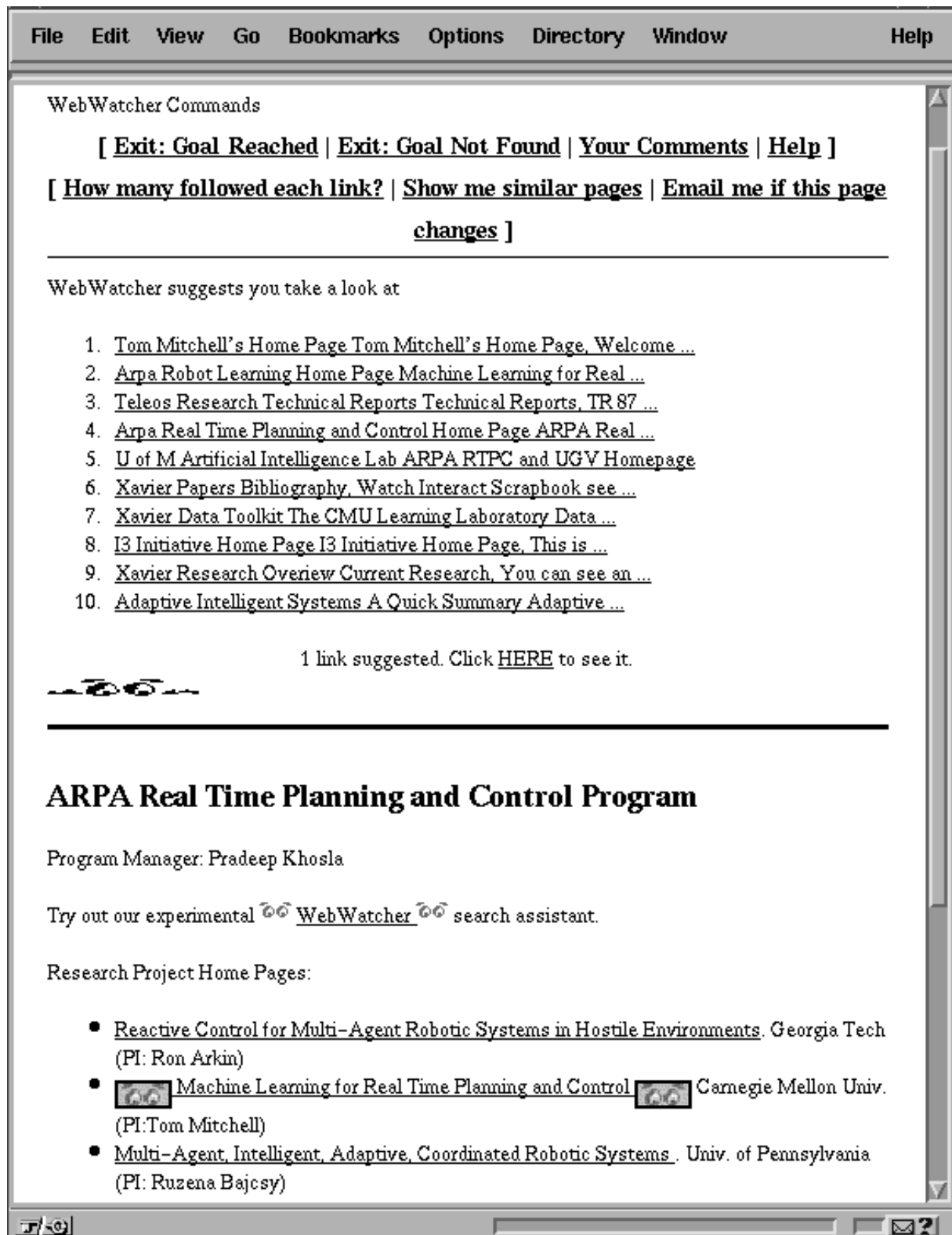


図 7: WebWatcher の動作の様子

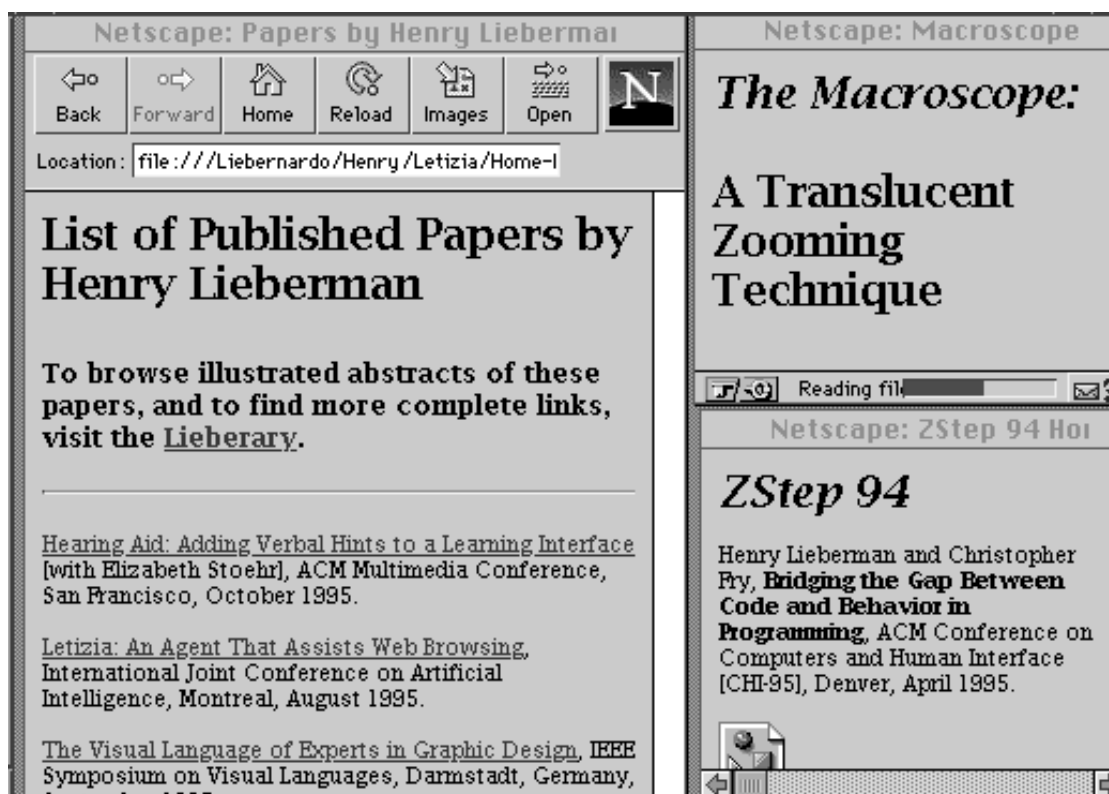


図 8: Letizia の実行画面

は 30 単語からの特徴ベクトルで表す。実際にはこれら 3 つのものを一つにした特徴ベクトルを用いている。ユーザがひとつページを選ぶたびにこの特徴ベクトルがひとつつくられることになる。

この特徴ベクトルの学習の仕方として、TFIDF (term frequency times inverse document frequency) 法 [51] を含め 3 種類の学習方法を用いて実験を行なっている。結果としては 3 つの学習方法いずれでもランダムに比べると 2 倍から 3 倍の正しさをリンクを提案していると報告している。また、リンクをネットワークとみなした時の学習も行なっている [30]。

5.1.2 Letizia

Letizia [36] も Web Watcher と同様にブラウジングを通じて学習するシステムである。しかし、Web Watcher とは異なり、特定のサーバを必要とせず、完全なクライアントとして動作する。

Letizia は現在ユーザが参照しているページを起点に推奨ページを提案する。図 8 の 3 つの画面のうち、左はユーザが自分でブラウジングをしているウィンドウであり、右下は Letizia が現在ユーザに推奨するページである。ユーザは通常左の画面でブラウジングしているが、右下の推奨ページが気に入れば、この画面に移ってブラウジングを続けることができる。

Letizia はユーザが現在参照中のページにあるリンクから幅優先で探索する。探索は一定数に達するかユーザが別のページに移動した時点で終了する。このため計算資源やネットワーク資源を

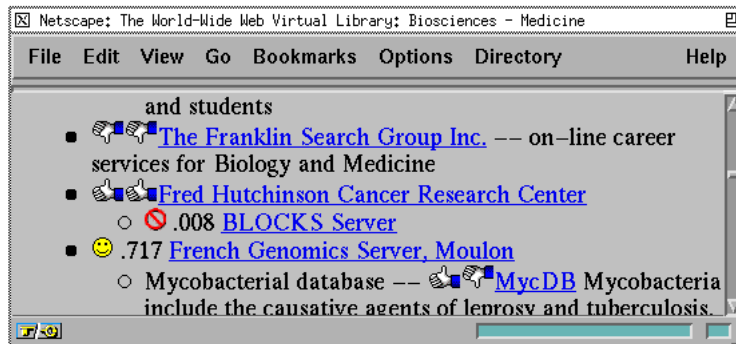


図 9: Syskill & Webert の画面

いたずらに消費することを避けている。ページの関連性の表現には TFIDF 法を用いている。また、探索の結果はユーザプロファイルとして保存され、次回以降でも継続的に用いられる。

5.1.3 Syskill & Webert

Syskill & Webert は同様にユーザの教示によって学習するブラウザである。ユーザはシステムに各ページが興味がある (hot) か、興味がない (cold)、どうでもよい (lukewarm) が指定する。その結果、これからまだみていないページに hot (上向きの指)、cold (下向きの指)、見た方がよい (スマイル)、見ない方がよい (ストップ) などを示す (図 9 参照)。また、学習した内容を利用して検索のためのキーワードを生成して、Lycos を使ってページを探してすることもできる。

このシステムが他の WebWatcher や Letiza と異なる点は、キーワードベクトルの元になるキーワードを分野をカバーするインデックスページを元にキーワード集合を自動生成する点である。与えられたページの中から意味ある単語を選び出すために次の式を用いている。

$$E(W, S) = I(S) - [P(W = present)I(S_{w=present}) + P(W = absent)I(S_{w=absent})]$$

ただし、

$$I(S) = \sum_{c \in \{hot, cold\}} -p(S_c) \log_2(P(S_c))$$

である。 $P(W = present)$ はそのページにある確率、 $I(S_{w=present})$ は W が一回でも出現するドキュメントの集合、 S_c はそのクラスに属するページを示す。

実験としては、Bayesian classifier、ID3、nearest neighbor アルゴリズム、perceptron、バックパゲーション付きの多層ニューラルネットワークを試しており、Bayesian classifier が分野によって偏りが少なく成績がよいと報告している。

1. n 個のエージェントにランダムな m ウェイトを与えて準備する。
2. エージェントが存在する限り以下を繰り返す。
 - (a) 今いるページの中の各リンク i について、質問単語の出現回数に対して単語の出現位置とリンクの間にあるヘッダの数で逆に重みづけしたものを L_i とする。そして、 $P_i = \exp(\beta L_i) / \sum_j \exp(\beta L_j)$ なる確率分布によって次の進むべきリンクを決定する。
 - (b) 新しいエネルギー状態を $E \leftarrow E - N + eR$ と計算する。 N はサーバアクセスコスト、 R は今いるページの質問単語出現回数。もし $(R + 1)\gamma$ が $1/2$ より大ならば、ユーザフィードバックエネルギー F により、 $E \leftarrow E + F$
 - (c) もし E が閾値 t を越えるならば、子供を生成する。子供のエネルギーは親のエネルギーを分配し、 β と γ はミュテートする。 E が負ならば、消滅させる。

図 10: 適応的情報エージェントのアルゴリズム

5.1.4 人工生命的方法の利用

Menczer ら [43][44] は人間のブラウジングに相当する行為を自己発生型の人工生命の方法を適用した適応型情報エージェントによって行なった例を報告している。エージェントはリンクをたどっていき、そのページの内容によってエネルギーを得たり、場合によっては増殖していく。アルゴリズムの概略を図 10 に示す。ここで β は最良優先探索 ($\beta = \infty$) からランダム探索 ($\beta = 0$) までの探索戦略を決定するパラメータである。 γ はユーザにそのページを提示するかどうかを決定するパラメータで、もしユーザが同意すればボーナスエネルギーがもらえる。

結果としては、エージェントは答のない質問の場合は急速に消滅し、それ以外の場合はすばやく答のページにエージェントは分布するようになったと報告している。またユーザのフィードバックが効果的であることも示している。

6 情報統合

情報の統合としては、情報分類、情報抽出、情報組織化などが挙げられる。分類は収集してきた情報を適当なカテゴリーに分けることであり、クラスタリングなどの方法がこれまで行なわれてきた。抽出とは収集してきた情報から必要な情報を抜き出すことであり自然言語理解などに関連する。組織化とはさらにそれらの情報の関係づけを行なうことであり、発想支援などとの関わりがある。ネットワークを用いた知的情報統合では、これら情報統合が情報収集と独立に行なわれるのではなく、お互いに影響しあって行なわれるところに特色がある。また、情報統合を自分が行なうのではなく、他のところで行なったものを利用する場合も考えられる。

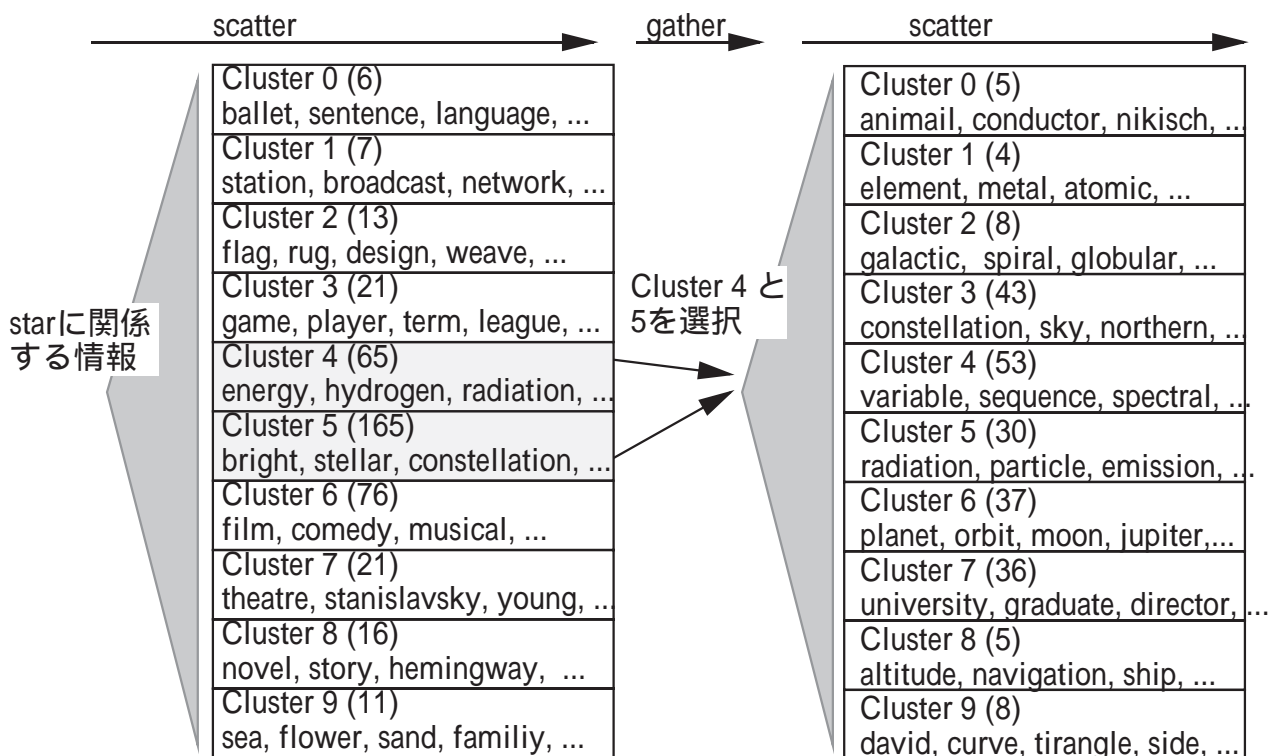


図 11: Scatter/Gather の分類の例

6.1 情報分類

情報分類の方法としては、予め分類方法を決めておく方法と与えられた情報の集合から分類項目を生成する方法がある。後者の場合、いかに処理時間を許容範囲に押えるかが問題となる。

6.1.1 Scatter/Gather

Scatter/Gather[25][26] は検索結果などの多量な情報を対話的にクラスタリングするシステムである。Scatter/Gather では入力された情報の集合を 10 のクラスタに自動分類し、各クラスタには話題になっている単語と典型的な表題を付加して表示する。各クラスタの話題単語はそのクラスタで頻出するキーワードであり、典型的表題はクラスタの重心に近い要素の表題である。ユーザはその興味ある任意個のクラスタを選択すると、今度はその選択されたクラスタに分類された情報のみを対象にして再び 10 のクラスタを生成する。これを繰り返すことにより、ユーザは漠然とした興味からでも自分の得たい情報にたどり着くことができる (図 11 参照)。

このシステムでは繰り返しクラスタリングを行なうので処理時間は重要である。このために階層的クラスタリング (hierarchical clustering) ではなく、分割型クラスタリング (partitional clustering) を用いている [12]。基本的にはまず、クラスタの種となる要素を決め、他の要素はそれぞれが最も近い種のクラスタに分類されるというものである。

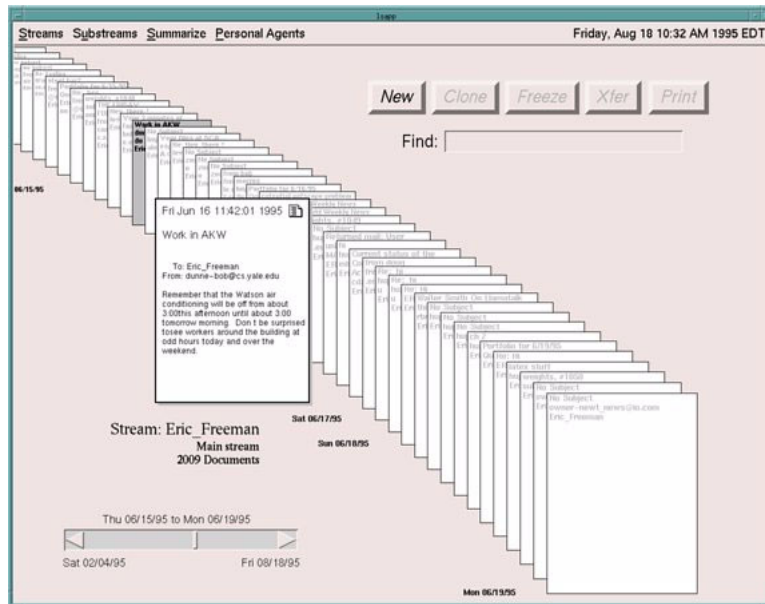


図 12: Lifestreams の概観

6.1.2 GALOIS/ULYSSES

GALOIS/ULYSSES[10]では束 (lattice) を用いて情報を分類表示する方法を示している。このシステムではキーワードの有無で束を作り、この束の上でユーザは情報をブラウジングする。また、ユーザは質問という形で束の一部を切り出したり、制約をつけて探索空間を狭めて束を作り直すこともできる。

6.1.3 Lifestreams

Lifestreams[18]はこれまで述べてきたような情報の分類や組織化を行なわず、時間系列に着目して、個人の電子情報を管理するシステムである。システムの概観は図 12に示すようになっており、情報はストリームと呼ばれる時間を軸とした列に並べられている。この時間軸にそってスクロールすることでブラウジングを行なう。そしていつかの操作 (New, Clone, Xfer, Find and Summary) が実行可能である。ユーザはサブストリームを作ることもできる。時間軸を中心とするアイデアは野口 [49] の考えと類似する。

6.2 情報抽出

情報抽出 (Information Extraction) は文章から情報を取り出すと意味では、自然言語処理であるが、いわゆる文章理解を行なうのではなく、必要な情報だけを効率的に取り出す点に重点をおいている [11]。ネットワーク上の情報のように多種多様な情報がある場合、正確な理解よりも、ユーザが欲する情報があるかないか、あるとすれば何かといった粗い理解が必要である。このためには情報抽出の考え方は重要である。

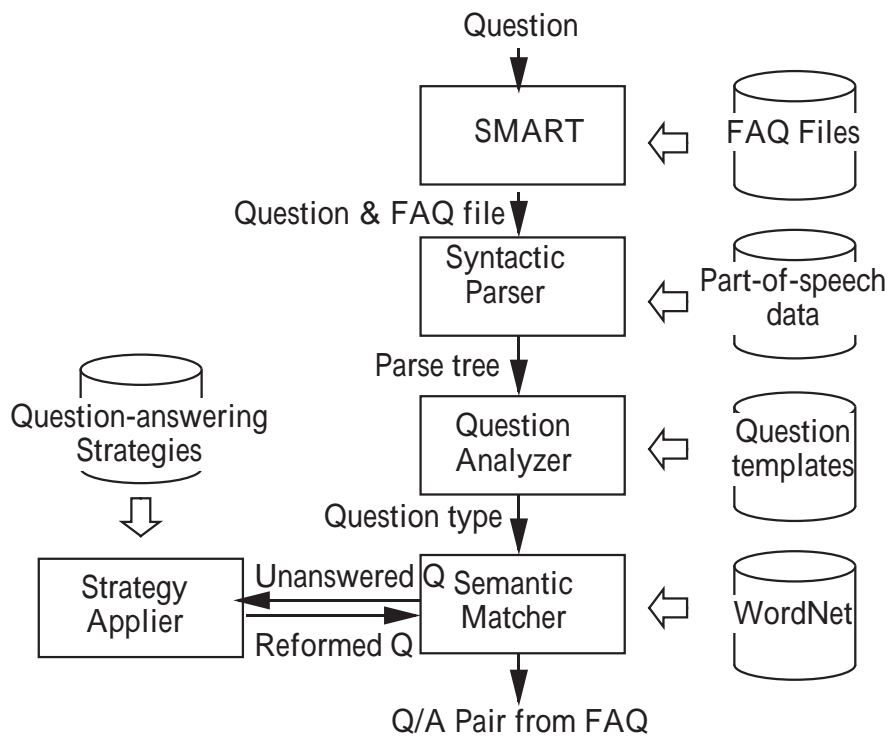


図 13: FAQ Finder の処理の流れ

情報抽出の方法としては事前に抜き出す項目が決定してる場合とそうでない場合がある。佐藤らの電子ニュースのダイジェスト自動作成や IICA では抜き出す項目が事前に決定されている。この方法は予め情報の構造がある程度仮定できる時に効率性と正確さが期待されるので、分野を限定した場合など有効である。これに対して、住田らの自動抄録生成や FAQ Finder は抽出される部分は比較的自由度が高く、より幅広い情報を入力して許している。

6.2.1 FAQ Finder

FAQ Finder[9][22][23] は直接 WWW のような大規模な情報源を利用しないが、多様な情報源と多様な利用者の結び付けを現状の人工知能の方法でどの程度実現できるか示しているよい例である。FAQ Finder はユーザの任意の質問に対して、収集してある FAQ (Frequently-Asked Questions) 情報³から適切な質問回答の組を選びだして、提示するというものである。その処理の流れを図 13に示す。

1. 質問に適切な FAQ ファイルを選び出すために SMART[7] を用いて統計的な情報検索を行なう。

³ネットワークニュースで頻繁に質問される質問を質問とその回答という形でまとめたもの。ニュースグループや話題毎に自発的に作られている。

2. 構文解析を行ない、単純な構文解析木をつくる。これにより主要な動詞と名詞句を同定する。
3. 質問の型を同定する。システムには予め質問形の分類がされており、構文解析木をこの中の一つに当てはめる。
4. WordNet[45] の意味ネットワークを用いて、意味的な概念マッピングを行なう。
5. もし近い質問が見つからなかった場合、質問回答戦略を用いて質問を変形させる。戦略としては比較質問を個別の質問への分解、「いつ」を「どうやって」に変更、などがある。

ここでの処理の重要な点は必ずしも深い言語理解を行なわない点である。構文解析は行なうが、主要な要素を見出すためのみに用いている。意味理解は浅い語彙の意味のマッチングしか行なわない。FAQ Finder の場合、出力は Q/A の組を選択することであり、回答を生成するわけではないので、これで十分である。

6.2.2 電子ニュースのダイジェスト自動作成

佐藤らの電子ニュースのダイジェスト自動作成 [52] では電子ニュースに投稿される会議告知記事あるいは論文募集記事からタイトル、開催期日、開催場所などを抽出するシステムを作成している。抽出の方法は、形態素解析は行なわず、文章のスタイル(センタリング、箇条書など)と言語パターン(「お知らせ」、「開催」、「x 月 y 日」など文字列パターン)およびそれらの順序情報を利用して行なっている。図 14 では、会議広告記事から各会議の開催日、会議名、開催場所が抜き出されて表示されている。結果としては、51%から78%の正解率であったと報告している。

6.2.3 住田らの文書構造解析に基づく自動抄録生成

住田らの文書構造解析に基づく自動抄録生成 [55] では書式(章節などの階層関係)と修辞構造(文と文の修辞的關係、例えば「順接」「例示」など)を元に、与えられた文章から重要文を抜きだし、再構成して提示するシステムを提案している。ここでは文の重要度の判定は修辞構造から行なっている。最重要文 74%、重要文 51%の捕捉率であったと報告している。

6.2.4 METIS

METIS は HTML 化した金属材料論文を要約とサーベイを自動的に行なうシステムである [42][41]。材料系の論文に特有な試験材料、実験方法、実験結果などの項目それぞれを抽出内容と抽出手段と組み合わせた定型的なスキーマで表現し、このスキーマを適用することで論文の重要な情報を

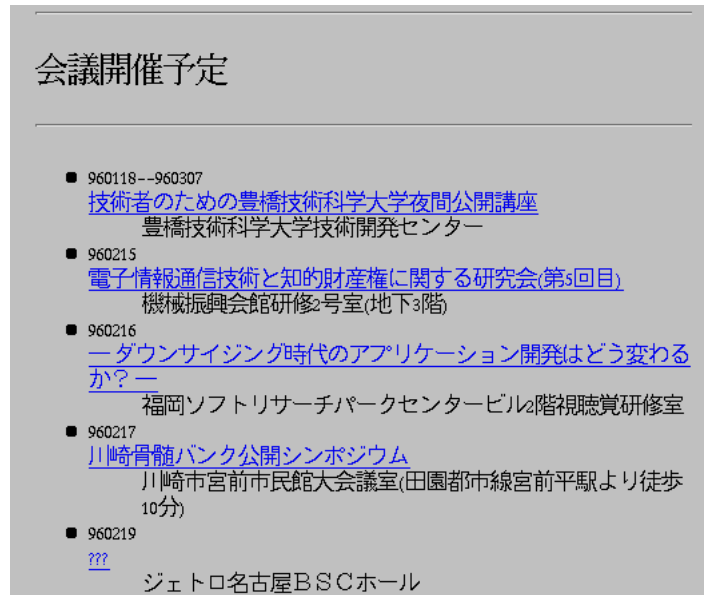


図 14: 電子ニュースのダイジェスト自動作成の結果の例抽出する (図 15)。抽出としては、キーワードによる文の選択、言い回しによる文の一部の切り出しなどを行なっている。

6.2.5 IICA

IICA は WWW から情報収集・分類・抽出を行なうシステムである [28][29]。このシステムの特徴はオントロジーと呼ばれる概念体系を情報の収集・分類・抽出の 3 つの段階全てを通して利用することである。このシステムの特徴は収集、分類、抽出の 3 つの段階において、オントロジーというかたちで記述された知識表現を使うことである。ここでのオントロジーは

6.3 情報組織化

情報分類や情報抽出は情報の量的な問題を解決することには貢献するものの、質的な向上は難しい。このためにはバラバラであった情報から関連性を積極的に見い出していく情報組織化といった作業が必要である。情報組織化では、いかに情報間の共通性を見い出すことができるかということと、その共通性をどう表現していくかということが問題になる。共通性発見に対する方法としては、情報のなかの個別要素の共通性に着目する、あるいは大域的な共通性に注目する方法 (例えば CAT1[54]) などの方法が考えられる。共通性表現の問題としては、以下で述べる二つのシステムのように情報のなかの概念を抽出して、その概念を中心に組織化を行なう方法がある。

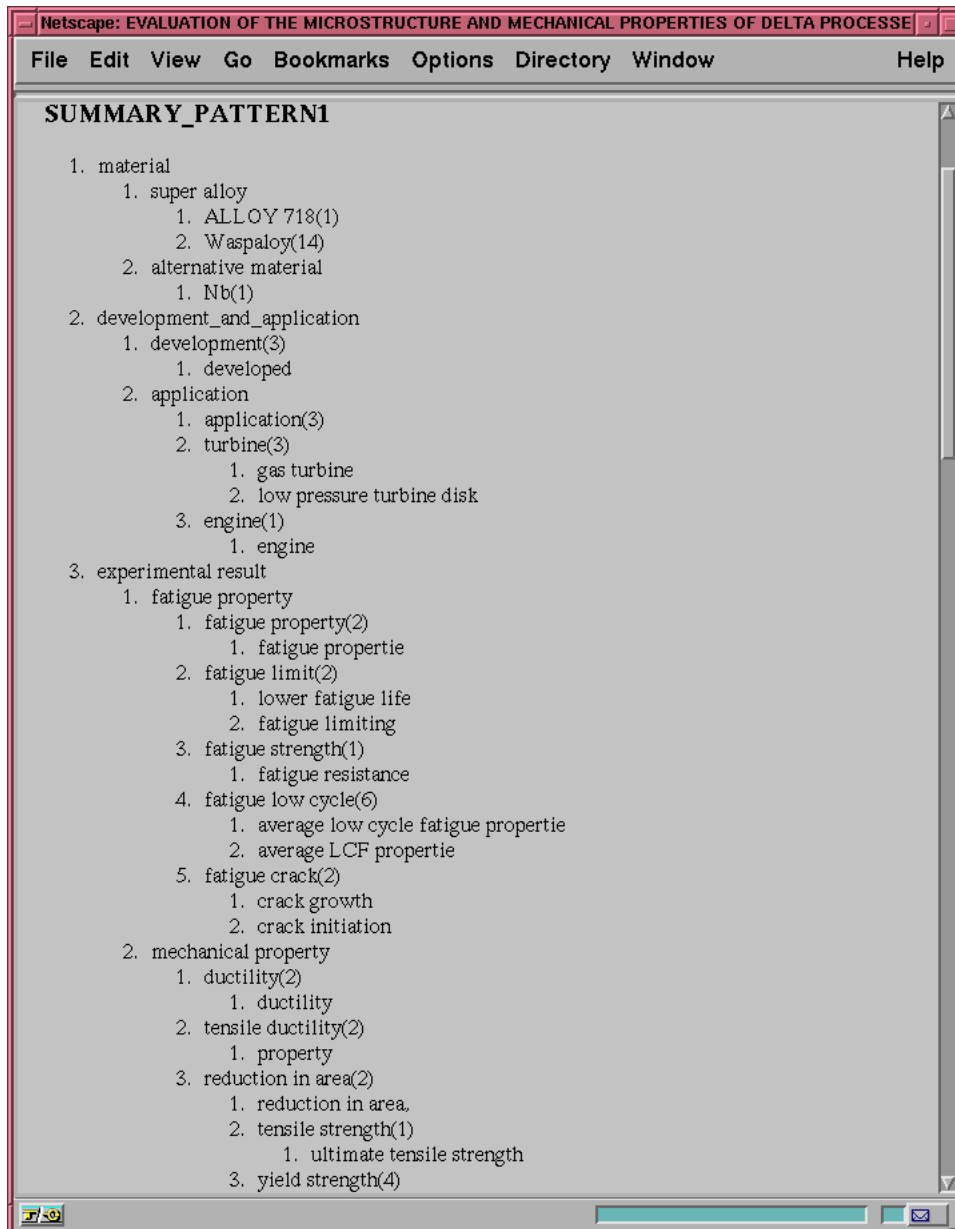


図 15: METIS の出力例

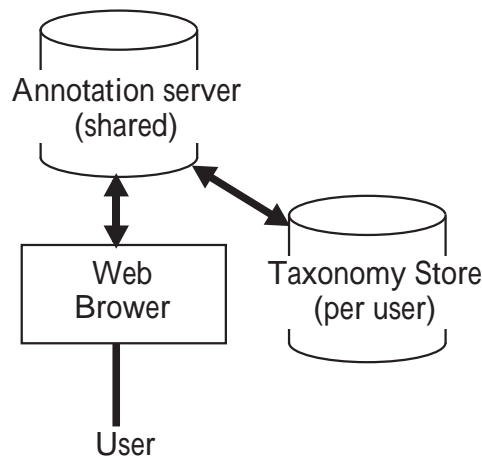


図 16: Active Notebook の構成

6.3.1 Active Notebook

Active Notebook[57] は個人の情報分類・組織化とグループの情報共有を支援するシステムである。Active Notebook は WWW ブラウザとグループで共有する注釈サーバと個人毎にもつ分類ベースからなる。図 16 にその概略を示す。

対象とする情報は WWW の情報、それに対する注釈としての個人の情報、非 WWW の個人の情報である。非 WWW 情報はその手がかりとなる情報を HTML 化することで他の WWW 情報と同等に扱う。

各情報には表題と概念を個人で付加する。概念は構造化された分類 (taxonomy) として個人ベースで蓄えられる。構造化としては "is a kind of" などがあり、複合語は自動的にその要素の単語とこの関係で結ばれる。ユーザはある情報を見ている時に、その情報と同じ概念を持つ情報を検索することができる。

個人の分類は二つの方法で共有可能である。ひとつはグループなどの共通分類を用いる方法で、もう一つは個別に他の人の分類を採り入れることである。

6.3.2 CM-2

CM-2[33][38][47] は非構造的情報を人間のもつ構造化作用を積極的に引き出すことで、構造化を行なうことを目指したシステムである。このシステムの中では全ての情報は概念ユニットのそれらの間の単一の連想関係で表現される (図 17 参照)。この連想関係はユーザが自分で作成するか、WWW などの外部の情報から生成することができる。この連想関係はシステムの中に用意されたいくつかのアルゴリズムによって構造化が行なわれる。

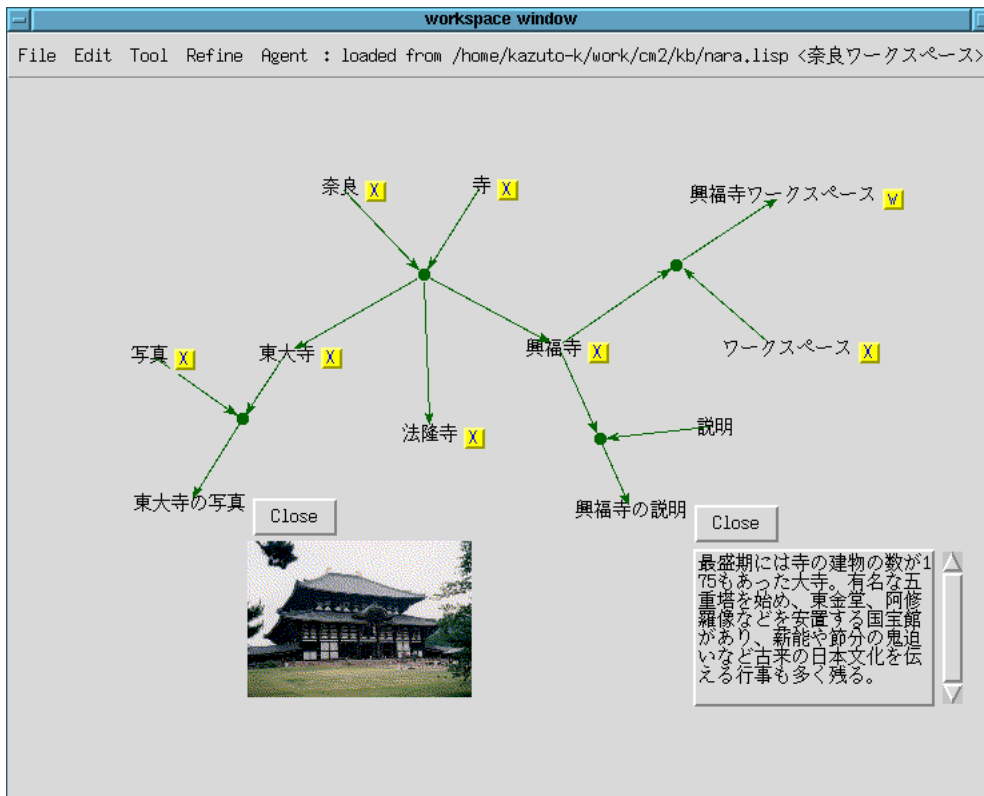


図 17: CM-2 の情報ベースの例

7 情報流通

情報流通では、情報源が他数ある場合、いかに情報源と情報利用者を結び付け、情報を伝達するかという問題を情報源間の関係として解決する立場の研究である。

ネットワーク上に存在する情報源の特徴としては、形式的多様性、内容的多様性、内容的分散性、形式的分散性などが挙げられる。

形式的多様性に対処するためにはどう情報源を均質化するかが問題となる。エージェント指向アプローチによる情報源のエージェント化がこの問題解決の一つの方向である。KQML[16]は知識レベルでのエージェント間コミュニケーションを目指したもので、エージェントのメッセージ送受の基本的な機能が規定されている。KQMLは例えば、information broker[15]、UMDL agent[5][58]、KC0 agent[48][56]などで用いられている。TSIMMIS[19]では、Wrapper生成器というものを用意して、データベースなどに均質なインタフェースを持たせる仕組みを用意している。

内容的多様性に対処するには情報源のモデル化をする必要がある。情報源のモデル化としては、オントロジーを用いる方法がある [21][56]。Information Manifold[35]やSIMS[1]ではdescription logicを元にした記述で表現するとしている。あるいはUMDLでは概要 (conspectus) というものを各情報源に持たせている。

内容的分散性から発生する問題としては、いかに情報源間の関係を見出すかことでくる。多くのシステムで仲介 (mediation)[59] という形で解決を試みている。TSIMMIS では関係を陽に記述して、そこから 仲介エージェント (mediator) を生成している。KC0 ではオントロジー上の関係から情報源間の関係を同定している。SIMS や Softbot[14] では論理的推論によって同定されるとしている。

さらに形式的多様性からは、複数の情報源からの情報獲得をどう実行するかということが問題となる。Softbot では観察と実行を論理的なプランニングで行なっている。SIMS ではプランニングと実行を統合することで効率良い情報獲得のプランができるとしている。

情報流通の研究の場合、上で述べた概念すべてがそれぞれに含まれることが多く、単純な分類は難しい。以下では、これらの概念に留意しながら、個々のシステム・研究を概観する。

7.1 データベース的アプローチ

ここでは、既存のデータベースなど比較的粒度の大きく、明確な情報源を対象としている。問題意識としては、いかに確実に仲介を行なえるかということが大きい。

7.1.1 TSIMMIS

TSIMMIS[19] は mediator アーキテクチャに基づく異種情報源の統合システムである。情報源としては既存のデータベースを利用しており、データベース型の比較的固定的なシステムである。全体のアーキテクチャを図 18 に示す。このシステムの中では情報は OEM (Object Exchange Model) というもので管理される (図 19)。そして、ここでは wrapper と呼ばれるプログラムが各情報源を隠蔽して、統一したインタフェースを与えている。Wrapper はテンプレートから生成される (図 20)。そして、各情報源に直接アクセスいく他、mediator を通じてアクセスすることができる。Mediator は OEM のよる統合的表現を使い、複数の情報源からの情報を統合した情報を新たにつくり出すことできる。この mediator は予め決められた mediator 仕様に基づいて動作する。図 21 では二つの文献データベースからの情報を統合する例である。

このように TSIMMIS での mediator は仕様の明確でかつ変化の少ない情報源を対象に、明示的な関係を見い出せるときに有効であるが、情報源が変化したり、情報源間の関係が自明でないときは、困難になる。

7.1.2 UMDL Agents

ミシガン大学デジタル図書館 (UMDL) プロジェクトでは図書情報に限らず、さまざまなネットワーク情報を利用できるような枠組みを用意している [5][58]。各情報には話題や形式、想定利用者、変更頻度などの概要 (conspectus) をつける (図 22 参照)。情報源側には CIA エージェントという情報源の概要を元に外部に均質なインタフェースを提供する殻となるエージェント、ユー

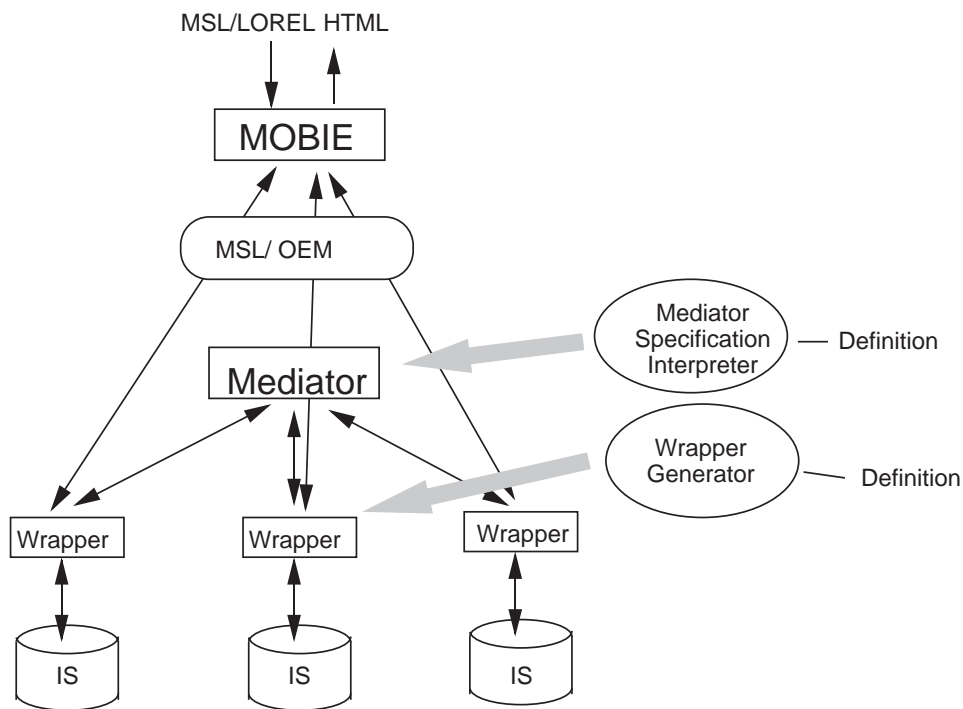


図 18: TSIMMIS のアーキテクチャ

OEM object

```

<inspectentry {
  <author "Hideaki Takeda">
  <title "Network-enhanced ...">
  <abstract "In this paper ...">
  <keywords "WWW, Information Gathering, ...">
}
>

```

Query

```

<answer {<author A><title T>} :-
  <inspectentry {<author A><title T>} AND
  author(A, "Takeda").

```

図 19: OEM の例

```

Inspec Rule
B:- B:<&O inspecentry {<author A><title T><abstract B>}> AND author(O, $X)

Inspec Action
// "find author $X" //.

```

図 20: Wrapper テンプレートの例

```

<&f(I,S) join_entry {<title T><authors A><abstract B><year Y>}>
  :- <&I inspecentry {<title T><abstract B>}>@inspec AND
  <&S bibliography {<title T><authors A><year Y>}>@biblio.

join_authors(f(I,S), W)
  :- written_by(I, W)@inspec AND author(S,W)@biblio

```

図 21: mediator 仕様の例

ザ側には UIA エージェントというユーザには独自のインタフェースと外部には同じく均質なインタフェースを持つエージェントがあり、その間を仲介エージェントが取り持っている。

同様な情報源の均質化の方向としては、InfoHarnes[53] ではより粒度の細かい方法で行なっている。

7.2 Ontology 指向アプローチ

このアプローチでは内容的な多様性をいかに処理できるかが課題である。この他 Stanford 大学では電子商取引の問題まで含めて研究を行なっている [15]。

7.2.1 Ontology に基づく仲介と変換

武田ら [56] は多重オントロジーに基づく仲介を提案している。これは情報源をオントロジーを用いて表現する方法であるが、オントロジーに多重性があるため、質問と情報源のオントロジーが異なる場合、オントロジーの変換が必要となる。オントロジーは Ontolingua[20] を拡張した ASPECTOL という言語でアスペクトという単位ごとに記述される。アスペクトに情報源そのものを指す基本アスペクトと情報源の統合を示す組合せアスペクト、情報源間の選択的關係を示すカテゴリ・アスペクトがある。カテゴリ・アスペクトにおいては複数の要素アスペクト間で同一視や変換可能な論理式の関係が記述される。

Collection Title	Blue Skies
Description	Recipient of one of Apple Computer “cool Tools Award for Internet Programming Efforts”, this ...
Content Provider	Weather Underground, AOSS Department, University of Michigan
Right Owner	Weather Underground, AOSS Department, University of Michigan
Conspectus Contact	James Alloway, cerebus@engin.umich.edu
Broad Topic(s)	Earth sciences
Specific Topic(s)	Atmospheric science
Searchable Form(s)	
Deliverable Form(s)	Text: HTML (Level 3), Software, Macintosh/Windows
Audience	Students: K-12, Teachers: K-12, Adult: General, Juvenile: General
Treatment	Introductory, Guide
Language	English
Chronological Coverage	1995
Update Frequency	Daily
URL	http://cirrus.sprl.umich.edu/
Native Search Engine	

図 22: UMDL の Conspectus の例 [5]

情報源 1 : 1990 年からの日本車

$v_1(model, year, price, owner) : CarFS(model, year, price, owner) \wedge JapaneseCar(model) \wedge year \geq 1990.$

情報源 2 : ラグジュアリー車

$v_2(model, year, price, owner) : CarFS(model, year, price, owner) \wedge price \geq \$20,000.$

情報源 3 : 外国車批評

$v_3(product, year, review) : Review(product, year, review) \wedge ForeignCar(product).$

情報源 4 : 低年式車批評

$v_4(product, year, review) : Review(product, year, review) \wedge Car(product) \wedge year \leq 1989.$

情報源 5 : 最近 5 年分の批評

$v_5(product, year, review) : Review(product, year, review) \wedge Car(product) \wedge 1991 \leq year \leq 1996.$

図 23: world-view による情報源の記述の例

7.3 論理的プランニング指向アプローチ

プランニングの手法を用いることで、いかに柔軟かつ動的な情報収集が可能になるかがこのアプローチの中心的課題である。

7.3.1 Information Manifold

Levy ら [35] は Information Manifold というアーキテクチャを提唱している。このアーキテクチャではより詳細な情報源の内容の表現とそれらからの情報収集のプランニングに焦点が当てられている。

このシステムの中では、情報源と質問は description logic の系統である CLASSIC[6] の記述に大小関係の述語を加えたもので表現される。CLASSIC ではオブジェクトのクラスを表現する 1 項述語 (concept と呼ばれる)、オブジェクト間の関係を表下する 2 項述語 (role と呼ばれる) ものからなる。これらの記述からクラス間の包含関係などを推論することができる。まず、情報源は world-view と呼ばれる、予め準備した class と role で統一的に記述される。図 23 は車のその批評に関するいくつかの情報源の表現である。ここで、 $CarFS(model, year, price, owner)$ や $Review(product, year, review)$ が world-view における role であり、 $JapaneseCar(model)$ や $Car(model)$ などが class である。クラスは階層的に包含関係を定義されている。

このとき、例えば「1988 年以降の 5000 ドル以下の日本車とその批評が欲しい」という質問は $q(Mo, Ye, Ow, Pr, Re) :$

$CarFs(Mo, ye, Pr, Ow) \wedge Review(Mo, Ye, Re) \wedge JapaneseCar(Mo) \wedge Pr \leq \$5000 \wedge Ye \geq 1998$

```

procedure generate-relevant-sources( $\mathcal{I} \ Q$ )
/*  $\mathcal{I}$  は情報源集合  $\mathcal{Q}$  は  $q(\bar{X}) : (\exists g_1(\bar{X}_1) \wedge \dots \wedge g_m(\bar{X}_m) \wedge C_q, C_q$  は  $\mathcal{Q}$  中の大小関係述語 */

for すべての subgoal  $g_i(\bar{X}_i), 1 \leq i \leq m$  に対して do:

     $rs_i = \emptyset$ 

    for  $\mathcal{I}$  中の情報源記述  $(v, r_v)$  の  $r_v$  中の大小関係述語以外の要素  $u(\bar{Y})$  に対して do:

        if  $g_i = u$  または  $u$  が非選言クラスならば then

             $\phi$  を以下の条件満たす  $r_v$  上での写像とする

                if  $Y$  が  $\bar{Y}$  の  $j$  番目の変数であり、 $r_v$  で存在限量子で束縛されていない

                    then  $\phi(Y) = X_j$  ( $X$  の  $j$  番目の変数  $X_j$ )

                    else  $\phi(Y)$  は新規の変数

                 $C(Q)$  を  $Q$  の制約、 $C(v)$  を  $\phi(r_v)$  の制約とすると、

                if  $C(Q) \wedge C(v)$  が充足可能ならば  $\phi(r_v)$  を  $rs_i$  に加える

```

図 24: 関連情報源の決定アルゴリズム

のように表現される。

この質問に答えるには先にあげた情報源を組み合わせる必要がある。意味のある組合せをつくり出すのが、ここでのプランニングということになる。

プランニングは2つのフェーズからなる。まず、質問の各項に対して、関係ある情報源を数えあげる。次にその組合せが論理的に健全であるかをチェックする。最初のフェーズのアルゴリズムを図24に示す。先の例では最初のサブゴール $CarFs(Mo, ye, Pr, Ow)$ から情報源1を、2番目のサブゴール $Review(Mo, Ye, Re)$ から情報源3、4、5を集めてくる。

情報源1と3の組合せの場合、

$$\begin{aligned}
 & CarFs(Mo, ye, Pr, Ow) \wedge JapaneseCar(Mo) \wedge Review(Mo, Ye, Re) \wedge ForeignCar(Mo) \wedge \\
 & \wedge Ye \geq 1990 \wedge Pr \leq \$5000 \\
 \Rightarrow & CarFs(Mo, ye, Pr, Ow) \wedge JapaneseCar(Mo) \wedge Review(Mo, Ye, Re) \wedge Ye \geq 1998 \wedge Pr \leq \\
 & \$5000
 \end{aligned}$$

という式が成り立つので健全である。

このアプローチでは先の TSIMMIS に比べ、(1) 情報源を質問の形態に合わせて表現しないのでより自然な情報源の表現になる、(2) 固定的な変換ではなく論理的な推論による結び付けなので情報源の数や内容の変化に容易に対応可能などの利点がある。しかし、world-view を構築する困難や、その記述の限界などの問題がある。類似のアプローチとしては前向き推論でプランニングを行なう Occam[34] がある。


```

(available output sims
  (retrieve (?port-name)
    (:and (seaport ?sport)
      (port-name ?sport ?port-name)
      (channel-of ?sport ?channel)
      (channel-depth ?channel ?depth)
      (transport-ship ?ship)
      (vehicle-type-name ?shop "breakbulk")
      (max-draft ?ship ?draft)
      (< ?draft ?depth))))
    (define (operator join)
      :parameters (?join-op ?data ?data-a ?data-b)
      :precondition
      (:and (join-partition ?data ?join-op
        ?data-a ?data-b)
        (available local sims ?data-a)
        (available local sims ?data-ab))
      :effect (available local sims ?data))
  )
)

```

図 25: 質問の表現例

図 26: join 操作の定義

7.3.2 SIMS

SIMS[1] はデータベース、知識ベースなどを統合して扱えるような情報メディエータを研究しているプロジェクトである。SIMS では情報は description logic の言語である LOOM[37] をもちいており、基本的には概念 (concept) とその役割 (role) で記述される。この SIMS のなかのシステム Sage[32] は分散情報源における情報収集のためのプランナーである。分散環境では、情報源がアクセス不能になったり、新しいゴールが途中ではいつてきたりする。Sage はプランニングと実行を密に結合して、必要に応じて、プランニングと実行を同時あるいは交互に行なうことで、上記のような分散環境下でのプランニングを可能にしている。

Sage では基本的な操作として、データ集合に対する操作 move, join, データ集合と情報源を結び付ける操作 select-source, データに対する操作 select, compute, assignment, 質問に対する操作 generalate, specialize, definition, decompose などを用意している。図 25 に質問の表現例を示す。図 26 に join の定義を示す。

Sage は基本的には、ゴールが与えられるとまず、完全なプランを生成し、それから実行する。プランは部分順序性を持つので、並列実行可能な部分は並列に実行される。しかし、実行中に部分プランの失敗や他のプランとの干渉がおきうるので、これを常時チェックして、必要な場合再プランニングを行なう。プラン中の各アクションは未実行、実行中、完了、失敗のいずれかにマークされる。実行中があるのは分散環境では必ずしも一つのアクションが短時間で終ると予想できないためである。プランニングや実行の必要性がある部分を flaw と呼び、これは未展開の条件あるいは干渉の可能性 (プランニング)、未実行と実行中 (実行) が含まれる。プランニングと実行のアルゴリズムを図 27 に示す。この例として図 28 に Sage の実行中の状態の例を示す。この例では、二つのゴール (port-name、port-description) を解こうとしている。最初の

現在のプラン集合から一つプランを抜き、以下のことを適用する。

- もし干渉の可能性があるならば、適当な制約をこのプランに加えて解決せよ。必要があれば、プラン集合に修正を加えよ。
- もし未展開の条件があるならば、アクションやリンクを加えて解決せよ。必要があれば、プラン集合に修正を加えよ。
- もし実行中のアクションが完了したら、
 - 実行が成功ならば、結果を記録してプランを更新せよ。プランが完了ならば、結果を出力して終了。それ以外なら、この新しいプランをプラン集合に加える。
 - 実行が失敗ならば、失敗したプランの部分を抜くと共に、失敗の再現しないためにモデルを更新する。この変更したプランをプラン集合に加える。
- 新しいゴールが加わったら、未展開の条件に加えて、このプランをプラン集合に加える。
- もし未実行アクションが実行可能になったらならば、これを実行するプロセスを生成する。更新したプランをプラン集合に加える。

図 27: Sage のプランニングと実行のアルゴリズム

ゴールのプランの一部が失敗して、再プランニングを行なった。このときもうひとつのゴールのプランのうちの実行中のアクションが、生成したアクションと同じ情報源を使っていることがわかったため、順序の制約をつけている。

7.3.3 Softbot

Softbot[14] はarchie や gopher など様々な情報源を組み合わせてユーザの希望する情報を動的にみつけるシステムである。Softbot はどういう情報源を組み合わせてゴールを達成するかをプランニングする。例えば、名前を元に E-mail アドレスを知りたい場合、まず文献データベースを引いて所属などの情報を得る。この情報を元に netfind コマンド使って E-mail アドレスを獲得する。プランニングの元になる知識 (スキーマ) は別に用意してあり変更可能である。また存在限量子のついた質問や制約のついた質問など不完全なゴールをもちいてもプランニング可能である。

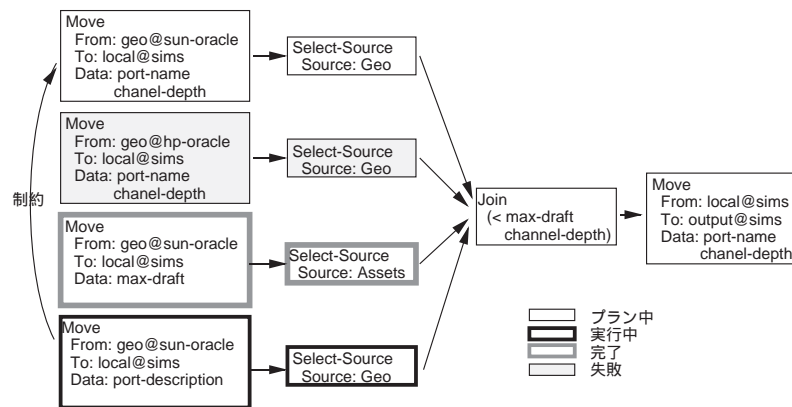


図 28: Sage の実行中の状態の例

8 まとめ

本稿ではネットワークを用いた知的情報統合に関わる研究を概観した。現在、様々なシステムが様々な方法を用いて構築されている。これは使う側がどういう視点、どういう用途を期待しているかによって、用いられる技術・手法も異なってくるからである。これまで情報源の多様性に関しては多くの考察がなされているが、使う側の多様性についての考察はまだ不十分である。使う側の多様性を意識することで、ネットワーク情報を利用した情報システムの課題は明確化し、より発展するものと思われる。

9 関連文献

関連の会議としては、AAAIの一連のシンポジウムが興味深い。1994年の春のシンポジウム”Software Agents”[13], 1995年春のシンポジウム”Information Gathering from Heterogeneous, Distributed Environments”[31], 秋のシンポジウム”AI Applications in Knowledge Navigation and Retrieval”[8], 1996年春のシンポジウム”Machine Learning in Information Access”[24]がある。また、AAAI-96併設のワークショップ”Internet-Based Information Systems”[17]もある。国内では、人工知能学会ホットトピックスと並列人工知能研究会の第2回研究会「特集テーマ：インターネット情報システム」[27]が関連がある。

また、本稿で言及したあるいは本稿に関連ある論文、システムなどでWWW上からアクセスできるものは<http://ai-www.aist-nara.ac.jp/~takeda/iii/>にまとめて公開している(付録として本稿にも掲載)。

10 謝辞

本稿をまとめるにあたり、奈良先端科学技術大学院大学の西田豊明教授からの示唆および議論が有意義でありました。心より感謝致します。

参考文献

- [1] Y. Arens, C. Y. Chee, C.-N. Hsu, and C. A. Knoblock. Retrieving and integrating data from multiple information sources. *International Journal of Intelligent and Cooperative Information Systems*, 2(2):27–158, 1993.
- [2] R. Armstrong, D. Freitag, T. Joachims, and T. Mitchell. Webwatcher: A learning apprentice for the world wide web. In *1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pp. 6–12, 1995.
- [3] M. Balabanović and Y. Shoham. Learning information retrieval agents: Experiments with automated browsing. In *1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pp. 13–18, 1995.
- [4] R. Basili and M. T. Pazienza. Semantic indexing of document bases. In *1995 AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 1–7, 1995.
- [5] W. P. Birmingham, E. H. Durfee, T. Mullen, and M. Weillman. The distributed agent architecture of the University of Michigan Digital Library (extended abstract). In *1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pp. 19–24, 1995.
- [6] R. Brachman, A. Borgida, D. McGuinness, P. Patel-Schneider, and L. Resnick. Living with classic: When and how to use a kl-one-like language. In J. Sowa, editor, *Principles of Semantic Networks*, pp. 401–456. Morgan Kaufmann, San Mateo, CA, 1991.
- [7] C. Buckley. Implementation of the SMART information retrieval system. Technical Report TR 85-686, Department of Computer Science, Cornell University, Ithaca, New York, May 1985.
- [8] R. Burke. AI Applications in Knowledge Navigation and Retrieval: Papers from the 1995 AAAI Fall Symposium. Technical Report FS-95-03, AAAI Press, March 1994.
- [9] R. Burke, K. Hammond, and J. Kozlovsky. Knowledge-based information retrieval from semi-structured text. In *1995 AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 15–19, 1995.
- [10] C. Carpineto and G. Romano. A system for conceptual structuring and hybrid navigation of text databases. In *1995 AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 20–25, 1995.
- [11] J. Cowie and W. Lehnert. Information extraction. *Communications of the ACM*, 39(1):80–91, 1996.

- [12] D. R. Cutting, D. R. Karger, J. O. Pederson, and J. W. Tukey. Scatter/Gather: A cluster-based approach to browsing large document collections. In *15th Annual International ACM/SIGIR Conference*, pp. 318–329, 1992.
- [13] O. Etzioni. Sofeware Agents: Papers form the 1994 AAI Spring Symposium. Technical Report SS-94-03, AAI Press, March 1994.
- [14] O. Etzioni and D. Weld. A softbot-based interface to the internet. *the Communications of ACM*, 37(7):72, 76 1994.
- [15] R. Fikes, A. Farquhar, and W. Pratt. Information brokers: Gathering information from heterogeneous information sources. In J. H. Stewman, editor, *Ninth Florida Artificial Intelligence Research Symposium (FLAIRS-96)*, pp. 192–197, 1996.
- [16] T. Finin, D. McKay, R. Fritzson, and R. McEntire. KQML: An information and knowledge exchange protocol. In K. Fuchi and T. Yokoi, editors, *Knowledge Building and Knowledge Sharing*. Ohmsha and IOS Press, 1994.
- [17] A. Franz and H. Kitano, editors. *Internet-Based Information Systems: Papers from the 1996 AAI Workshop*. AAI Press, August 1996.
- [18] E. Freeman and S. Fertig. Lifestreams: Organizing your electronic life. In *1995 AAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 38–44, 1995.
- [19] H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman, and J. Widom. Integrating and accessing heterogeneous information sources in tsimmi. In *AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Envrionments*, pp. 61–64, 1995.
- [20] T. R. Gruber. Ontolingua: A mechanism to support portable ontologies. Technical Report KSL 91-66, Stanford University, Knowledge Systems Laboratory, 1992.
- [21] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. Technical Report KSL 93-04, Knowledge Systems Laboratory, Stanford University, August 1993.
- [22] K. Hammond, R. Burke, and S. L. Lytinen. A case-based approach to knowledge navigation. In *Proceedings of IJCAI-95*, pp. 2071–2072, 1995.
- [23] K. Hammond, R. Burke, and C. Martin. FAQ Finder: A case-based approach to knowledge navigation. In *AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Envrionments*, pp. 69–73, 1995.
- [24] M. A. Hearst and H. Hirsh. Machine Learning in Information Access: Papers from the 1996 AAI Spring Symposium. Technical Report SS-96-05, AAI Press, March 1996.
- [25] M. A. Hearst, D. R. Karger, and J. O. Pederson. Scatter/Gather as a tool for navigation of retrieval results. In *1995 AAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 65–71, 1995.
- [26] M. A. Hearst and J. O. Pederson. Revealing collection structure through information access interfaces. In *Proceedings of IJCAI-95*, pp. 2047–2048, 1995.

- [27] ホットトピックスと並列人工知能研究会 (第 2 回). Technical Report SIG-HOT/PPAI-9602, 人工知能学会, 1996.
- [28] 岩爪, 白神, 畑谷, 武田, 西田. テキストからの情報抽出・統合化法の提案と知的情報収集・分析システム IICA の実験的評価. 第 7 回データ工学ワークショップ, 1996.
- [29] M. Iwazume, H. Takeda, and T. Nishida. Ontology-based information capturing from the internet. In *Proceedings of the fourth International Conference on the International Society of Knowledge Organization*, pp. 261–272, 1996.
- [30] T. Joachims, T. Mitchell, D. Freitag, and R. Armstrong. Webwatcher: Machine learning and hypertext. In *Fachgruppentreffen Maschinelles Lernen*, Dortmund, August 1995.
- [31] C. Knoblock and A. Levy. Information Gathering from Heterogeneous, Distributed Environments: Papers from the 1995 AAI Spring Symposium. Technical Report SS-95-08, AAAI Press, March 1995.
- [32] C. A. Knoblock. Planning, executing, sensing, and replanning for information gathering. In *Proceedings of IJCAI95*, Montreal, Canada, 1995.
- [33] 糞谷, 西田. 乱雑な情報ベースを組織化するための情報ベース構築手法. 人工知能学会研究会資料 SIG-KBS-9503, pp. 59–64, 1996.
- [34] C. T. Kwok and D. S. Weld. Planning to gather information. In *Proceedings of AAAI96*, pp. 32–39, 1996.
- [35] A. Y. Levy, A. Rajaraman, and J. J. Ordille. Query answering algorithms for information agents. In *Proceedings of AAAI96*, pp. 40–47, 1996.
- [36] H. Lieberman. Letizia: An agent that assists web browsing. In *Proceedings of IJCAI-95*, pp. 924–929, 1995.
- [37] R. MacGregor. A deductive pattern matcher. In *Proceedings of AAAI88*, St. Paul, MN, 1988.
- [38] 前田, 西田. 知識メディアシステム CM-2 とそのユーザインタフェース. 第 11 回ヒューマンインタフェースシンポジウム論文集, pp. 49–54, 1995.
- [39] P. Maes. Agents that reduce work and information overload. *Communications of the ACM*, 37(7), 1994.
- [40] P. Maes and R. Kozierok. Learning interface agents. In *Proceedings of AAAI-93*, pp. 459–465, 1993.
- [41] 松尾, 武田, 西田. KP 化による論文内容の効果的提示方法とその応用. ヒューマン・インタフェース・シンポジウム, pp. 581–588, 1995.
- [42] T. Matsuo and T. Nishida. Intelligent support for construction and exploration of advanced technological information space. In *Proceedings of the fourth International Conference on the International Society of Knowledge Organization*, 1996. (To appear).

- [43] F. Menczer and R. K. Belew. Artificial life applied to adaptive information agents. In *1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pp. 128–132, 1995.
- [44] F. Menczer, W. Willuhn, and R. K. Belew. An endogenous fitness paradigm for adaptive information agents. In *CIKM'94 Workshop on Intelligent Information Agents*, 1994.
- [45] G. A. Miller. WordNet: A lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- [46] 西田. ソフトウェアエージェント. *人工知能学会誌*, 10(5):704–711, 1995.
- [47] T. Nishida, K. Koujitani, and H. Takeda. A plain indexing method for organizing conceptually promiscuous data. In *1995 AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 103–109, 1995.
- [48] T. Nishida and H. Takeda. Towards the knowledgeable community. In *Proceedings of International Conference on Building and Sharing of Very Large-Scale Knowledge bases '93*, pp. 157–166, Tokyo, 1993. Japan Information Processing Development Center.
- [49] 野口. 「超」整理法. 中央公論社, 1993.
- [50] D. Oard. Information filtering resources, March 1996. <http://www.enee.umd.edu/medlab/filter/>.
- [51] G. Salton and M. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., 1983.
- [52] 佐藤, 佐藤, 篠田. 電子ニュースのダイジェスト自動作成. *情報処理学会論文誌*, 36(10):2371–2379, 1995.
- [53] L. Shklar, S. Thatte, H. Marcus, and A. Shet. The "InfoHarness" information integration platform. In *the Second World Wide Web Conference '94*, 1994.
- [54] 角, 堀, 大須賀. テキストオブジェクトを空間配置することによる思考支援システム. *人工知能学会誌*, 9(1):139–147, 1994.
- [55] 住田, 知野, 小野, 三池. 文書構造解析に基づく自動抄録生成と検索提示機能としての評価. *電子情報通信学会論文誌*, J78-D-II(3):511–518, 1995.
- [56] H. Takeda, K. Iino, and T. Nishida. Agent organization and communication with multiple ontologies. *International Journal of Cooperative Information Systems*, 4(4):321–337, December 1995.
- [57] M. C. Torrance. Active Notebook: A personal and group productivity tool for managing information. In *1995 AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 131–135, 1995.
- [58] M. P. Wellman, E. H. Durfee, and W. P. Birmingham. The digital library as community of information agents. *IEEE Expert*, June 1996. (to appear).
- [59] G. Wiederhold. Mediation in the architecture of future information systems. *IEEE Computer*, 25(3):38–49, 1992.