

ネットワークを利用した知的情報統合

Network Enhanced Intelligent Information Integration

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

Keywords: WWW, information gathering, information integration, information classification, information extraction, information organizing

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2100万の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。

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が、WWWでは多様かつ広範な領域の情報が存在し、利用法は自明ではない。このため、単に情報収集ができるだけでは不十分で、分類・抽出など情報の利用まで含めて考える必要がある。これを本稿ではネットワークを用いた知的情報統合と呼ぶことにする。

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1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、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) [Oard, 96]、ブラウジング (browsing) の方法が挙げられる。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。

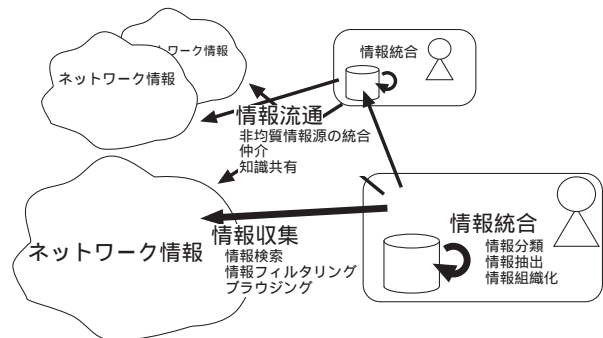


図1: ネットワークを利用した知的情報統合の枠組み

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。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。

また情報流通の問題もある。これ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。具体的には非均質情報源の統合、仲介、知識共有などがテーマとなる。

¹<http://www.altavista.digital.com/>

2 ネットワークを用いた知的情報統合を可能とする技術

前節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あるかといった概略の理解である。通常の構文解析を用いているものもあるが(例えば[Basili 95])、対象が非均質であるがために汎用性が問題になる。こ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[Burke 95][Hammond 95])や着目点の維持しつつ、いかにユーザの行為を支援できるかなどがある。

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。汎用的な概念体系を目指したものとしてはIICA[Iwazume 96]のオントロジーがある。また、語彙体系としてはWordNet[Miller 95]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な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[西田 95]。MaesのLearning Interface Agents[Maes 94][Maes 93]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。

前節で述べたようなネットワークを用いた知的情報統合の全ての要素を持つシステムを構築するのは容易でなく、実際にはそれぞれのシステムがこのなかのいろいろな側面に着目して構築されている。以下ではそういったシステムを概観する。なお、本稿では図1で挙げた要素のうち、ブラウジング、情報分類、情報抽出、情報組織化、非均質情報源の統合に着目する。最近の情報検索については住田ら[住田 96]や木谷[木谷 96]を参照されたい。また、知識共有に関しては西田[西田 94]を、仲介に関してはWiederhold[Wiederhold 92]を参照されたい。

3 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、ユーザの意図を維持しつつ、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3つの研究を取り上げる。他にもBalabanovićら[Balabanović 95]の研究もある。

3.1 Web Watcher

Web Watcher[Armstrong 95]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2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

²<http://www.yahoo.com/>

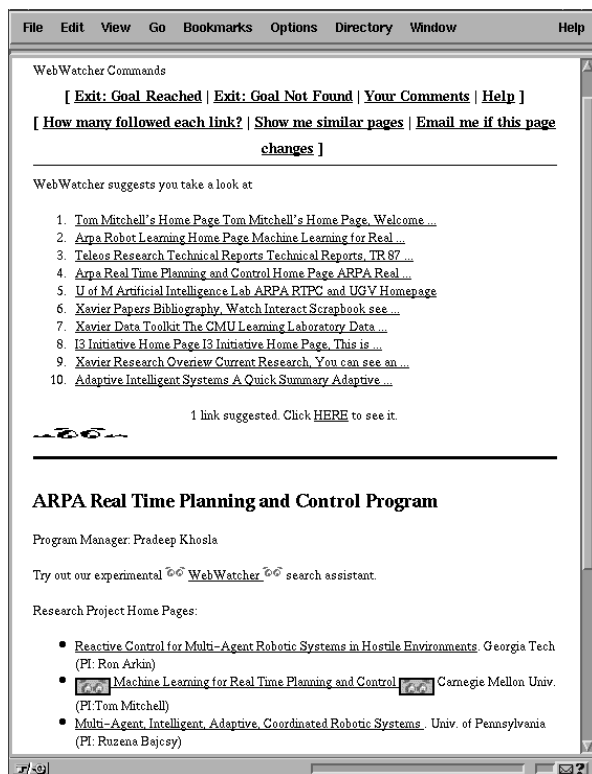


図 2: WebWatcher の動作の様子

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 [Salton 83] 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 [Salton 83] 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 倍の正しさで

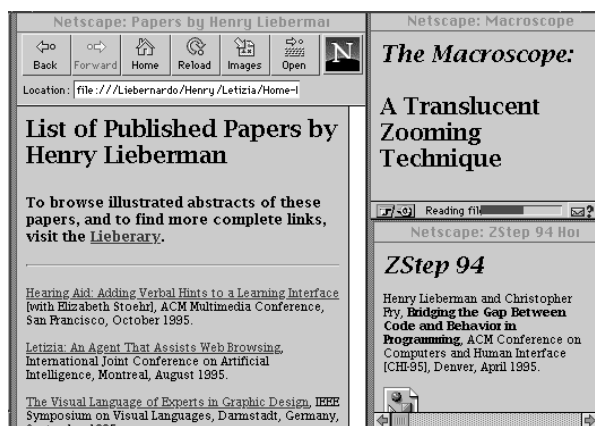


図 3: Letizia の実行画面

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 [Joachims 95]。

3.2 Letizia

Letizia [Lieberman 95] 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 3 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インドウであり、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ィールとして保存され、次回以降でも継続的に用いられる。

3.3 人工生命的方法の利用

Menczer ら [Menczer 95][Menczer 94] 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

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 はミューテートする。 E が負ならば、消滅させる。

図 4: 適応的情報エージェントのアルゴリズム

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。

4 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。

4.1 Scatter/Gather

Scatter/Gather[Hearst 95b][Hearst 95a] 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、典型的な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 5 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

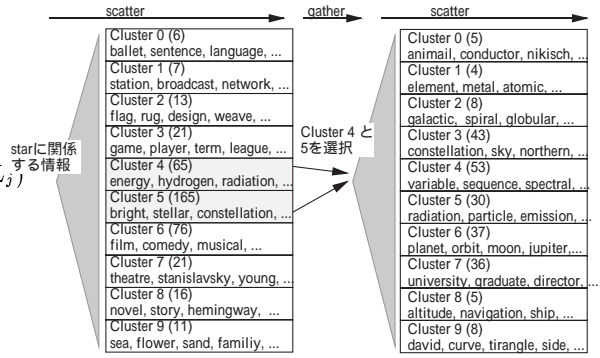


図 5: Scatter/Gather の分類の例

[Cutting 92]。基本的にはまず、クラスタの種類となる要素を決め、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。

4.2 GALOIS/ULYSSES

GALOIS/ULYSSES[Carpineto 95] 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。

4.3 Lifestreams

Lifestreams[Freeman 95] はこれまで述べてきたような情報の分類や組織化を行わず、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。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 [野口 93] の考えと類似する。

5 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 [Cowie 96]。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。

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いる場合とそうでない場合がある。佐藤らの電子

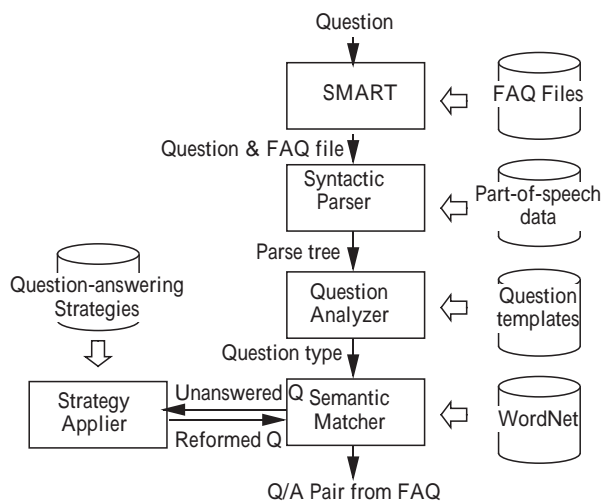


図 6: FAQ Finder の処理の流れ

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。

5.1 FAQ Finder

FAQ Finder[Burke 95][Hammond 95a][Hammond 95b] IICA は WWW から情報収集・分類・抽出を行なうシステムである [岩爪 96][Iwazume 96]。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。

1. 質問に適切な FAQ ファイルを選び出すために SMART[Buckley 85] を用いて統計的な情報検索を行なう。
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[Miller 95] 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。

³ ネットワークニュースで頻り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。

URL	温泉の名前	最寄り駅	アクセス方法	風呂の種類	泉質
akase-spa-j.html	赤瀬温泉		"バス"		"炭酸鉄泉"
hinagu-spa-j.html	日奈久温泉	"JR 八代駅"	"JR 日奈久駅下車"		"食塩泉" "単純"
kanaketa-spa-j.html	金桁温泉	"JR 三角駅"	"バス"		"炭酸鉄泉"
tsurugiyama-spa-j.html	鶴木山温泉	"JR 佐歌駅"			"単純"
tsuruyu-spa-j.html	鶴湯温泉		"徒歩"		"単純"
yoshio-spa-j.html	吉尾温泉	"JR 吉尾駅"	"徒歩"		"単純"
yanoko-spa-j.html	兎温泉	"JR 水俣駅"	"バス"	"沖合いの湯"	"重曹泉"
yanotsuru-spa-j.html	鶴温泉	"JR 水俣駅"	"バス"		"単純"
yanoura-spa-j.html	湯浦温泉	"JR 湯浦駅"	"徒歩"		"単純"

図 7: IICA での情報抽出例

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。

5.2 IICA

情報収集としては、オントロジー上の概念を指定すると、その概念および関連する概念に相当するキーワードでページ上のリンクを評価し、関連があると判定されたリンクをたどるという形でページを収集していく。情報分類としては、オントロジー上の概念に関連つけられたキーワードを元に特徴ベクトルを生成して、各ページを最も近い概念に配分していく。

情報抽出として状態遷移による方法と特徴記述ルールによる方法の 2 つの方法を用いている。後者はオントロジー上の概念やその属性それぞれに抽出のためのルールを付加して、それを用いて情報抽出を行なっている。抽出のためのルールは言語パターンであり、これにより必要な文を抜きだし、次に文の一部を切り出して概念の属性値を埋めていく。ルールはオントロジーを用いて構造化されているので、複雑な抽出ルールも概念の組合せで構成的に書くことができる。抽出した情報の例を図 7 に示す。

5.3 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 [佐藤 95] 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。この結果51%から78%の正解率であったと報告している。

5.4 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 [住田 95] 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。

5.5 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 [Matsuo 96][松尾 95]。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抽出する。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 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出し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[角 94])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。

6.1 Active Notebook

Active Notebook [Torrance 95] 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

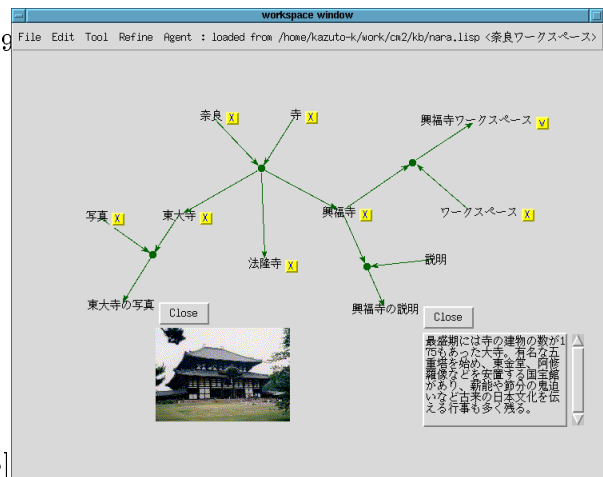


図 8: CM-2 の情報ベースの例

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。

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

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。

6.2 CM-2

CM-2 [糀谷 96][前田 95][Nishida 95] 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8参照)。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。

7 非均質情報源の統合

ネットワーク上に存在する情報はWWWだけでなく、多様なものがある。情報源が異なる形態・形式の場合、どのように統合するかが問題になる。その方法

としては情報源に一樣なアクセス情報をつけて情報源を均質化する方法と、問題に合わせて様々な情報源を組み合わせる方法であり、この二つは相補的に機能する。以下ではそれぞれの例となる研究を挙げる。

7.1 UDML Agents

ミシガン大学デジタル図書館 (UDML) 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 [Birmingham 95]。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。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 [Wiederhold 92] 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 [Shklar 94] ではより粒度の細かい方法で行なっている。

7.2 Softbot

Softbot [Etzioni 94] 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。

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 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

- [Armstrong 95] Armstrong, R, Freitag, D, Joachims, T, and Mitchell, T: WebWatcher: A Learning Apprentice for the World Wide Web, *1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pp. 6-12 (1995).
- [Balabanović 95] Balabanović, M and Shoham, Y: Learning Information Retrieval Agents: Experiments with Automated Browsing, *1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pp. 13-18 (1995).
- [Basili 95] Basili, R and Pazienza, M. T: Semantic Indexing of Document Bases, *1995 AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 1-7 (1995).
- [Birmingham 95] Birmingham, W. P, Durfee, E. H, Mullen, T, and Weillman, M: The Distributed Agent Architecture of the University of Michigan Digital Library (Extended Abstract), *1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pp. 19-24 (1995).
- [Buckley 85] Buckley, C: Implementation of the SMART Information Retrieval System, Technical Report TR 85-686, Department of Computer Science, Cornell University, Ithaca, New York (1985).
- [Burke 95] Burke, R, Hammond, K, and Kozlovsky, J: Knowledge-based Information Retrieval from Semi-Structured Text, *1995 AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 15-19 (1995).
- [Carpineto 95] Carpineto, C and Romano, G: A system for conceptual structuring and hybrid navigation of text databases, *1995 AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 20-25 (1995).
- [Cowie 96] Cowie, J and Lehnert, W: Information Extraction, *Communications of the ACM*, Vol. 39, No. 1, pp. 80-91 (1996).
- [Cutting 92] Cutting, D. R, Karger, D. R, Pederson, J. O, and Tukey, J. W: Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections, *15th Annual International ACM/SIGIR Conference*, pp. 318-329 (1992).
- [Etzioni 94] Etzioni, O and Weld, D: A Softbot-Based Interface to the Internet, *the Communications of ACM*, Vol. 37, No. 7, pp. 72 (1994).

- [Freeman 95] Freeman, E and Fertig, S: Lifestreams: Organizing your electronic Life, *1995 AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 38-44 (1995).
- [Hammond 95a] Hammond, K, Burke, R, and Lytinen, S. L: A Case-Based Approach to Knowledge Navigation, *Proceedings of IJCAI-95*, pp. 2071-2072 (1995).
- [Hammond 95b] Hammond, K, Burke, R, and Martin, C: FAQ Finder: A Case-Based Approach to Knowledge Navigation, *AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pp. 69-73 (1995).
- [Hearst 95a] Hearst, M. A and Pederson, J. O: Revealing Collection Structure through Information Access Interfaces, *Proceedings of IJCAI-95*, pp. 2047-2048 (1995).
- [Hearst 95b] Hearst, M. A, Karger, D. R, and Pederson, J. O: Scatter/Gather as a Tool for Navigation of Retrieval Results, *1995 AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 65-71 (1995).
- [Iwazume 96] Iwazume, M, Takeda, H, and Nishida, T: Ontology-Based Information Capturing from the Internet, *Proceedings of the fourth International Conference on the International Society of Knowledge Organization* (1996), (To appear).
- [岩爪 96] 岩爪道昭, 白神謙吾, 畑谷和右, 武田英明, 西田豊明: テキストからの情報抽出・統合化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).
- [Joachims 95] Joachims, T, Mitchell, T, Freitag, D, and Armstrong, R: WebWatcher: Machine Learning and Hypertext, *Fachgruppentreffen Maschinelles Lernen*, Dortmund (1995).
- [木谷 96] 木谷強: インターネットにおける検索技術, 情報処理学会第52(平成8年度前期)全国大会シンポジウム・インダストリアルセッション論文集, pp. 177-185 (1996).
- [糞谷 96] 糞谷和人, 西田豊明: 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).
- [Lieberman 95] Lieberman, H: Letizia: An Agent That Assists Web Browsing, *Proceedings of IJCAI-95*, pp. 924-929 (1995).
- [前田 95] 前田晴美, 西田豊明: 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).
- [Maes 93] Maes, P and Kozierok, R: Learning Interface Agents, *Proceedings of AAAI-93*, pp. 459-465 (1993).
- [Maes 94] Maes, P: Agents that Reduce Work and Information Overload, *Communications of the ACM*, Vol. 37, No. 7, (1994).
- [松尾 95] 松尾利行, 武田英明, 西田豊明: 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).
- [Matsuo 96] Matsuo, T and Nishida, T: Intelligent support for construction and exploration of advanced technological information space, *Proceedings of the fourth International Conference on the International Society of Knowledge Organization* (1996), (To appear).
- [Menczer 94] Menczer, F, Willuhn, W, and Belew, R. K: An Endogenous Fitness Paradigm for Adaptive Information Agents, *CIKM'94 Workshop on Intelligent Information Agents* (1994).
- [Menczer 95] Menczer, F and Belew, R. K: Artificial Life Applied to Adaptive Information Agents, *1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pp. 128-132 (1995).
- [Miller 95] Miller, G. A: WordNet: A Lexical Database for English, *Communications of the ACM*, Vol. 38, No. 11, pp. 39-41 (1995).
- [西田 94] 西田豊明: 協調型アーキテクチャによる知識の共有と再利用, 人工知能学会誌, Vol. 9, No. 1, pp. 23-28 (1994).
- [Nishida 95] Nishida, T, Koujitani, K, and Takeda, H: A Plain Indexing Method for Organizing Conceptually Promiscuous Data, *1995 AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 103-109 (1995).
- [西田 95] 西田豊明: ソフトウェアエージェント, 人工知能学会誌, Vol. 10, No. 5, pp. 704-711 (1995).
- [野口 93] 野口悠紀雄: 「超」整理法, 中央公論社 (1993).
- [Oard 96] Oard, D: Information Filtering Resources (1996), <http://www.enee.umd.edu/medlab/filter/>.
- [Salton 83] Salton, G and McGill, M: *Introduction to Modern Information Retrieval*, McGraw-Hill, Inc. (1983).
- [佐藤 95] 佐藤円, 佐藤理史, 篠田陽一: 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, Vol. 36, No. 10, pp. 2371-2379 (1995).
- [Shklar 94] Shklar, L, Thatte, S, Marcus, H, and Shet, A: The "InfoHarness" Information Integration Platform, *the Second World Wide Web Conference '94* (1994).
- [角 94] 角康之, 堀浩一, 大須賀節雄: 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, Vol. 9, No. 1, pp. 139-147 (1994).

- [住田 95] 住田一男, 知野哲朗, 小野顕司, 三池誠司:
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,
Vol. J78-D-II, No. 3, pp. 511-518 (1995).
- [住田 96] 住田一男, 三池誠司: 知的情報検索の動向,
人工知能学会誌, Vol. 11, No. 1, pp. 10-16
(1996).
- [Torrance 95] Torrance, M. C: Active Notebook: A
Personal and Group Productivity Tool for Man-
aging Information, *1995 AAAI Fall Symposium
on AI Applications in Knowledge Navigation
and Retrieval*, pp. 131-135 (1995).
- [Wiederhold 92] Wiederhold, G: Meidation in the
architecture of future information systems,
IEEE Computer, Vol. 25, No. 3, pp. 38-49
(1992).

著者紹介



武田英明 (正会員)

1986年東京大学工学部精密機械工学科卒業。1988年東京大学大学院工学系研究科修士課程修了。1991年東京大学大学院工学系研究科博士課程修了。工学博士。同年財団法人日本システム開発研究所入所。1992年ノルウエー工科大学 (Norwegian Institute of Technologoy, Norges Tekniske Hoegskole, NTH) ポストドクトラル研究員。1993年奈良先端科学技術大学院大学情報科学研究科 助手。1995年同 助教授。1992年～1993年ノルウエー王立科学技術研究会議 (Royal Norwegian Council for Scientific and Industrial Research, Norges Teknisk-Naturvitenskapelige Forskningsraad, NTNF) research fellow。1995年人工知能学会全国大会優秀論文賞授賞。知識ベースシステム、知的CAD、設計学などの研究に従事。