

概念ブラウザと個人情報環境

藤澤浩道* 山崎直子** 橋本哲也**

* (株) 日立製作所 ソフトウェア開発本部

** (株) 日立製作所 中央研究所

個人情報環境は計算機を用いた情報活動を支援するところのソフトウェア環境である。その中で特に重要であると考えられる情報の分類、蓄積、検索を支援する個人用情報ベース「概念ブラウザ」について、その設計思想および概要を述べる。情報の蓄積・検索は情報検索技術とよく結び付けられるが、ここではむしろ、情報の整理・分類のためのツールが重要であることを論述する。意味ネットワーク型の知識ベースを用いて、蓄積情報に関する重要な概念を整理、分類し、文書などのメディア情報と関連づけて管理する。これにより、ユーザとシステムとが概念空間を共有することができ、上位概念からの検索や、概念空間におけるハイパーテキスト機能が可能となる。

Concept Browser and Personal Information Environment

Hiromichi FUJISAWA*, Naoko YAMAZAKI** and Tetsuya HASHIMOTO**

*Software Development Center, Hitachi, Ltd.

**Central Research Laboratory, Hitachi, Ltd.

Personal Information Environment is a software system to support various information activities of knowledge workers. Classification, storage and retrieval are important activities that should be supported by such environment. Presented in this paper is a personal information base, named "ConceptBrowser," and its design philosophy. The importance of organization and classification of incoming information is stressed. A semantic network type knowledge base is used to store important concepts in an organized manner. Documents and other kind of media information are stored in connection with the concepts kept in the knowledge base. This approach enables both the users and the system to share the "conceptual space," and in the sequel, the system can provide intelligent retrieval and associative browsing functions.

1. はじめに

個人情報環境は、ワークステーションなどを介して研究者等の知的ワーカーが諸々の情報活動を行なう際のソフトウェア環境をいう。この情報活動には、情報の収集、蓄積、分析、検索、利用・加工、通信・対話、議論、意思決定、計画などがある。本来、個人情報環境はこれらすべてを含むべき概念であるが、ここでは特に情報の蓄積・検索について議論する。

情報を機械に蓄積して必要に応じて即座に検索できるようにしたいという願望は古くからあり、1945年にはV. Bushが個人用文献システムMEMEX ("Memory Extender")を構想したことはよく知られている¹⁾。しかしながら、個人が集めた情報を整理したり、あるいは自分が理解した結果を忘れないように機械に蓄積・管理させ、我々の記憶能力を拡張するようなシステムを作るには、現在なお課題がある。

書類をキャビネットに仕舞うときに一抹の不安を感じるのはよくあることである。例えば、同一の案件に関係している複数の書類も、分類上の理由から別のところに格納すれば、その関連性は見えなくなってしまう。あるいは、分類体系を忘れて新たな分類項目を設定して仕舞ったり、さらには、分類体系を変更すべきことが分かっているにもかかわらずその作業が大変なためにその場しのぎの分類を行なって、結果として混乱することがある。

断片情報の整理と蓄積にも困っている。多くの場合、入ってくる情報は断片的かつランダムである。ノートやメモに残しても、散逸してしまう。本来、断片的あるいは不完全な情報を体系的に積み上げていくことは、どの分野の専門家にとっても重要なことである。

本論文では、このような問題を解決することを狙って研究を進めてきている個人用情報ベース「概念ブラウザ」の設計思想と概要、およびその有効性について述べる。

2. 情報検索と文書ファイリング

記憶に留めたい情報は多くの場合、文書の形態で入って来たり、あるいは文書の形態で保存する。したがって、情報の整理・蓄積・検索の問題は、従来の言葉を用いれば「文書ファイリ

ング」である。この観点で見ると、電子ファイリングシステムが計算機を用いた手段として存在するが、この問題を解決していない。その理由は、電子ファイリングシステムが大容量記憶装置に単に情報検索機能を付加した概念に留まっていることにある。

情報検索と文書ファイリングとは、いくつかの本質的な点で異なる。まず第一に、情報検索は膨大な情報（大きなデータベース）の中から、検索者にとっては未知の情報を検索するものである。したがって、ここでの問題は、「関連ありそうな文書」を如何に漏れなく検索するか、あるいは検索でヒットする情報を如何に少ない件数に絞り込むかにある。すなわち高い再現率と高い検索精度が求められる。

一方、文書ファイリングでは、検索するときには「あの文書」というように既知の情報を検索するのであり、その文書そのものが検索されねばならない。その意味では、文書ファイリングにおける検索の方が技術的に難しいといえる。ここでは、これを「同定型検索」と呼ぶことにする。

第二の相違点は、前者が文書を登録するのは別の管理者であるのに対して、後者は本人が後日その情報を取り出すことを意識しつつ行なう。したがって、文書ファイリングでは、利用者の記憶も重要な意味を持ち、これをも勘案したシステム設計が重要である。すなわち、このように利用者の記憶にも頼ることにより、実のところ、唯一に同定するところの同定型検索が可能になっている。

第三の相違点は、情報検索が蓄積内容が一定の範囲にあることを前提にしているのに対して、文書ファイリングでは対象となる文書とその分類（分類体系）は常に推移している。（実は、前者においても内容は時間とともに推移しているが、データベースの規模が大きいため追従することが困難であるのが実態であり、現在問題として顕在化している。）

このように文書ファイリングでは情報を如何に整理、分類しながら格納するかが問題である。

本研究の立場は、情報の整理・分類を計算機によって支援し、その結果として情報検索の問題にアプローチしようとするものである²⁾。

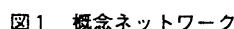
個々の情報の内容を客観的に、かつ簡単に記述できないことは、キーワード付けが難しいことなどでも証明されている。情報の整理・分類は、内容を記述する代わりに、他の情報との相対的な関係によって情報を表現する方法である。ある観点で類似している情報をグルーピングし、それらをさらに観点を変えて階層的に分類するのが普通である。このような階層的分類によって、後から、共通な属性が見えてきたりする。

分類法の利点は、分類体系を自由自在に改編できることである。一方、属性記述法の利点は、複雑な概念を個々には単純な属性の集合で表現できることである。例えば、分類法においては分類(グループ)の名称を任意の語句で表現すればよく、また、入れ替えをするだけで中身の制約を受けることなく分類が変更できるという融通性がある。一方、属性記述法では、細かな分類を行わずに、例えば、文書を作成者、作成日、タイトル等の属性で代表することができる。

は、読んだ結果を仕舞うことにはならない。読者（研究者）にとって重要なのは、例えば、その文献が述べている新しい概念であり、関連する自分の研究、あるいはその文書であろう。以下は、「情報の整理・分類」として計算機に登録したくなるであろうシナリオである。

- (R1) 文献を自分の大まかな分類体系に位置付ける。
- (R2) 著者名を含めた最低限の書誌事項を属性として登録する。
- (R3) もし著者が有名な研究者なら、別途、その研究者と研究機関についても登録する。また、もし発表会議が記憶に留めておくべき国際会議であればそれも別途登録する。
- (R4) 文献の最も重要な概念を語句で表現して新規登録するとともに、その語句で表した概念「X」を文献の「主題」属性として登録する。概念「X」は適切な位置に分類、登録する。
- (R5) もし、ある研究で利用できそうならば「関連研究」属性として登録する。

「UNIVERSAL」を根とする分類階層木（概念木と呼ぶ）の中に位置付けて記憶する。概念ノードには、一般概念と具体概念があり、前者は一般名称（総称）双方によって記述できる。属性が明確ではない抽象的な概念は分類によって管理し、類似した項目が数多くかつ属性が明確なものは分類とともに属性記述によって管理する。



これらの双方の利点を活かした情報整理・分類の支援は、意味ネットワーク型の知識表現方式を用いることにより実現できる。ここでは、任意の概念を表現したいので、「概念ネットワーク」と呼ぶことにする。上記の例では、「文献」、「研究者」、「研究機関」、「X」、「主題」、「研究テーマ」等は概念ノードを構成し、これらノードの下位には具体的な概念で代表されるものであり、後者は唯一に同定できるものである。すべての概念をここに網羅することはできないが、たとえばユーザとして研究者を想定した場合、およそ図2に示すような概念を含む。

人物、職種、職位、電話番号、E-mailアドレス、
組織・機関（政府機関、企業、教育機関、大学、
研究機関、学会、部、課、...）、
人工物（製品、プロトタイプ；ハード、ソフト、システム、
デバイス、装置、材料、...）、
地名（地方、国、都市、住所、...）、
時（年、年月、年月日）、
イベント（国際会議、委員会、学会、...）、
プロジェクト（国家プロジェクト、
研究プロジェクト、...）、
テーマ、目標、技術課題、
科学・技術（分野名、技術項目、具体的技術名）、
文書（書籍、雑誌、学会誌、論文、記事、議事録、
資料、...）、
文書ファイル、など

図2 登録した概念の一部

概念間に定義できる関係リンクには、概念の上下関係を表わすIs-A関係とPart-Of関係、およびそのほかの任意の関係がある。前者はいわば縦糸であるのに対して、後者は横糸に対応する。横糸に対応する意味的關係には、「研究者」と「論文」を結ぶ関係「発表」、あるいは「人」と「会議」を結ぶ関係「参加」などがある。これらの関係は、「研究者が発表した論文」とも、「論文を発表した研究者」とも表現できるように、双方向性のリンクで表わす。

実験では、約450件のコンピュータ関連の新聞記事を登録し、結果として約3,500の概念と約7,000の関係を登録した。

4. 人間の記憶特性と知的検索

検索の難しさは、一つには、明快に欲しいものを記述できない人間の特性に起因している。「あれ」とか「これ」といったあいまいな表現を使ったり、「このまえのあの時のあれ」といったイベントを持ち出す³⁾。

このような考察や、心理学的に知られていることをまとめると、以下の4点が、特に同定的検索へのアプローチにとって重要である。

(1) 記憶の抽象化

記憶は段々と薄れていくが、すべてを忘れるわけではない。事物の具体的な名称、たとえば固有名詞は忘れやすいが、それが何であったかは覚えている。事物を上位概念（上位語）で覚えていることが多い。したがって、このような上位概念からでも求める情報に到達できるようにすることが必要である。

(2) 関係の記憶（連想的記憶）

あるものを一つの単語で明確に表現できないときは、上位概念で説明的に記述することができる。たとえば、「医学情報を扱ったオンラインデータベース」とか、「データベース会社が在る都市」といった具合である。

また、事物と事物の間の連想性の重要性は、MEMEXの論文にも言及されており、現在のハイパーテキストの考え方の原点になっている。

(3) ランダムな記憶の想起

おぼろげな記憶を思い起こす過程はランダムである。記憶の鮮明な部分から思い出し、その思い出した記憶が刺激となってさらに新たなことを思い出す。たとえば、「AI構築ツール」が具体化され、「有名計算機会社が開発したAI構築ツール」となったり、「AI構築ツールを主題にした新聞記事」というように発展したりする。あるいは、「新聞記事ではなく雑誌記事だった」と言い換えたりしながら、段々と記憶が鮮明になっていく。そのプロセスは一様ではなく、システムに欲しいものを伝えるための柔軟な対話プロセスが不可欠である。

(4) 再刺激による記憶の想起

「想起と認識」の関係はよく知られている。覚えていることを陽に表現することは難しいが、

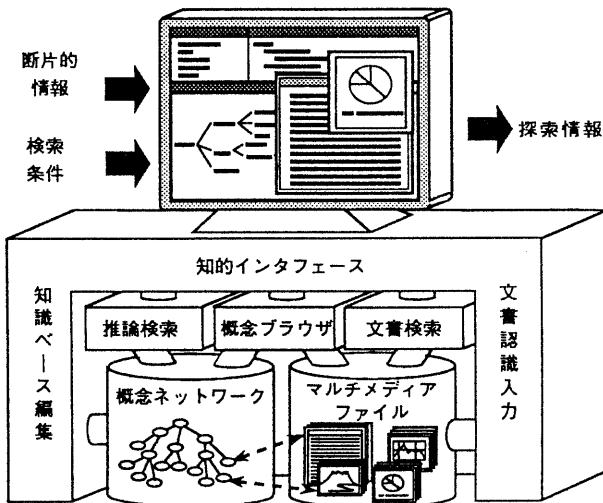


図3 個人用情報ベースのシステムイメージ

見せられたり言われたりすると「それだ」と認識できる。さらに、それが刺激となって新しいことを思い出す。

知的検索は、積極的に関連情報を表示し、ユーザが悩むことなく自然な対話によって、検索要求を入力できるようにするものである。

5. 個人用情報ベース

5.1 システムイメージ

本論文で議論する個人用情報ベースは、概念ネットワークと関連づけてメディアを蓄積・管理することが特徴である（図3）。すなわち、本システムは蓄積する情報に関する重要な概念を概念ネットワークの形で保有する。第3章「情報の整理・分類」で述べたように、情報を登録していく過程で、システムはユーザと共有する「概念空間」を概念ネットワークとして持つことになる。さらに、この共有する概念空間の存在によって、前述の知的検索が実現されることになる。

メディア情報（文章、図表、写真、音響など）は概念ノードと対応付けられていると同時に、相互の関連性はハイパーテキストリンクによって記憶される。

情報ベースとしての価値は、この概念ネットワークを随時、編集・更新できることにある。そのためのソフトウェアとして、「概念ブラウ

ザ」は

- 1) 概念と関係の登録・変更
- 2) 同義語・異表記語の登録・変更
- 3) 概念木の再構成
- 4) 事実文の登録
- 5) ブラウザと知的検索

の機能を提供する。情報の整理は、分類を随時見直ししながら、ものごとの体系を概念木に仕立てていくことである。時間的な断面を見ると、体系付けられ情報と、未分類の情報とが混在していてもよい。

概念ネットワークはある意味では常に未完成な知識であるといえる。名前が分からない人物を“unknown-person”として概念登録

してもよいし、「未分類」という概念ノードを作ってもよい。これらは分かった時点で更新すればよい。

文書の登録は、その文書を代表する概念ノードをまず概念ネットワークに新規登録し、同ノードに主題や作成者などの属性を関係リンクでつなぎ、その概念を定義することにより行なう。

5.2 拡張概念ネットワーク

概念ネットワークに複雑な事実を登録するには、図1に示す2項関係のモデルを多項関係モデルに拡張した。まず、「概念」を拡張して、これまでの「関係」をも概念のひとつとして位置づけ、「関係概念」として管理する（図4）。関係概念には、「～であること」を表わす「属性関係」と、「～すること」を表わす「動作関係」とがある。前者は2項関係で表わされる属性であり、後者は多項関係で表わす動詞概念である。従って、概念空間は物の世界、属性の世界、事象の世界を表現することができる。

関係の一般的枠組みは、格文法に基づくテンプレートで表現する。たとえば、動詞「製造する」のテンプレートは、主体は「メーカ」、対象は「製品」、場所は「工場」というように、概念「製造」の意味的な制約を規定する⁴⁾。具体的な事実文は、このテンプレートの穴埋めを行う要領で、たとえば、「日立は3050をK工場で製造する」と登録することができる。

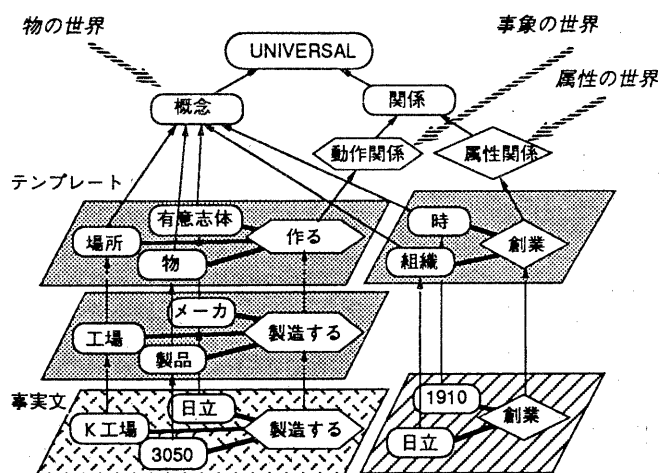


図4 拡張概念ネットワークによる知識表現

さらに、テンプレートには、蓄積情報を自然言語で表現するための文パターン、付加すべき助辞、動詞の活用形なども保持する。これにより、検索条件の表示や、蓄積している事実の表示を日本語や英語で行うことができる。

5.3 概念ブラウザ

概念ブラウザは、概念ネットワークを編集したり、概念を検索したりするためのソフトウェアである(図5)。

ユーザは階層木ウィンドウの概念木を見ながら、新規概念を登録したり、概念木を変更したり、さらには検索文を試行錯誤的に作成して検索できる。

ブラウジングに際しては、概念木を注目したい部分について詳細に広げたり、不要な部分を表示から削除したりすることが可能である。

検索文編集機能には、

- 1) 条件文追加・削除
- 2) 概念置き換え
- 3) 制約格追加・削除

があり、条件を弛めたり、狭めたりすることが容易に行なえる。検索文の鍵括弧内の概念は、概念木の中の一つの概念をピックアップすることで置換できる。図6は、「何か賞を受けたことがあるデザイナーが設計した建築物の資料」を意味的に検索した例である。

検索結果は文書概念であるが、それに付随する属性や事実文は、詳細情報として検索結果ブ

ラウザウィンドウに表示される。さらに、該当する文書イメージが表示される。

検索文は上位概念で説明的に記述した抽象概念であり、任意の抽象度でよいことが従来の検索方式と異なる点である。検索文中の概念は、テンプレートで規定されている範囲の任意の上位概念に置き換えることができる。「新聞記事」で狭すぎれば「記事」に置き換えることができるし、さらに「文書」に置き換えてもよい。

概念ブラウザは概念空間でのハイパーテキスト機能を実現する。表示されているすべての概念

はマウスで選択でき、それに関する情報を表示させることができる。図7は、動詞概念「受賞した」を選択して、システムが知っているすべての事象「受賞」を一覧表示した例である。当然、メディア情報がリンクされていれば、それを表示できる。

6. 有効性の検討と今後の課題

概念ネットワークエディタを用いて、有効性を示すため、コンピュータ関連の新聞記事と技術論文を合計約450件登録した。図2は、その結果構築されたネットワークの主な概念を示している。約3,500の概念と、これらを結ぶ約7,000の関係を登録した。このうち、約600の概念が一般概念で、もののクラスを表現するものであった。また、関係については、約90種類の意味的關係を登録した。

この過程で、主要な概念とそれらの間の一般関係は、新規情報の入力を容易にする上で極めて重要であることが分かった。概念ネットワークの基本部が構築された後は、1件の記事の登録には約5分を要した。記事に重要な概念を含んでいてそれを登録する場合には、10分掛かることもあった。

一方、基本部の定義には時間が長く掛かることが分かった。場合によっては、抽象的な概念を一つ登録するのに、関連する概念の登録が必要になり、30分以上の時間が掛かることもあった。主要な属性と関係で構成する概念ネットワーク基本部は、これをデータベースと見たとき

の「スキーマ」に対応し、事前のデータモデルの設計が重要であることを意味する。

個人用情報ベースの有効性は、登録した情報の価値と、そのために払うコストによって決まる。これらを定量的に評価することは難しいが、論文や記事を執筆する際に一つの情報を求めて極めて長い時間探し回ることもあり、情報ベースの必要性は一般的にも高いと思われる。

問題は、情報の入力コストを小さくすることである。本論文は、「個人用」として議論してきたが、グループでの知識の共有こそが重要であり、互いに集めた情報を共有し、入力コストを下げるのが求められる。

元来、MEMEXなども知識の蓄積を如何に共有するか、あるいは必要な人が必要な情報に如何に効率的にアクセスできるかが課題であり、グループウェアとしての「情報ベース」への展開が重要である。

7. まとめ

本論文では、概念ネットワークを用いた個人用情報ベースについて述べた。自分の手元に集まる情報を自分の価値観にしたがって体系的に個人データベース化するシステムが個人用情報ベースである。

最近個人でも文書処理で毎月1.5メガバイト

のフロッピーディスクを消耗することもまれではない。これを25年続けても、5インチの光磁気ディスク一枚に入る。12インチの光ディスク一枚には、厚さの20メートルの文書を画像として蓄積できる。

したがって、今後は、この情報を如何に組織化して、生きた情報と成すかが情報処理システムにとって課題である。個人用情報ベースは、ユーザとシステムが共有する概念空間を概念ネットワークとして記憶することにより、情報の組織化のニーズに応えることができると考える。

参考文献

- 1) F. W. ランカスター：「紙なし情報システム」，植村俊亮訳，共立出版，1984。
- 2) 藤澤浩道，畠山敦，木内伊都子，「情報整理ツールとしての知的ファイリングシステム」，昭和63年電気・情報関連学会連合大会，32-3，pp.5-97-5-100，1988。
- 3) 木内伊都子，畠山敦，大木優，藤澤浩道，「知的検索を目指したConcept Browser」，情報処理学会，情報学基礎研究会資料，13-4，1989.6。
- 4) J. F. Sowa, "Conceptual Structures: Information Processing in Mind and Machine," Addison-Wesley, 1984.

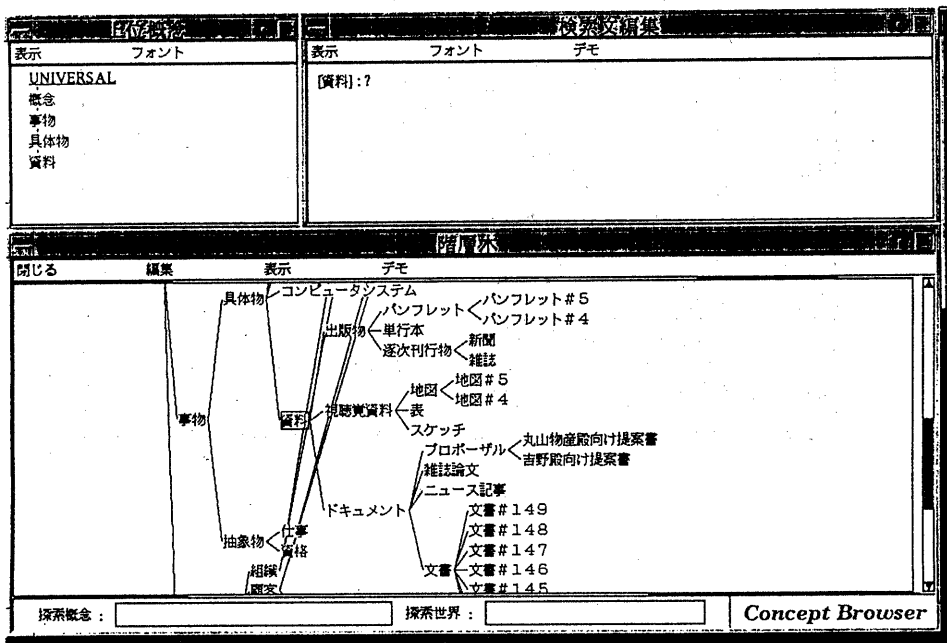


図5 概念ブラウザの画面



図6 検索結果の表示

受賞		登録	表示	閉じる
1	1975年に楠木かみしが毎日デザイン賞を受賞した。			
2	1987年にコリン・ゴールドがブリッツカニ賞を受賞した。			
3	1979年にエヴァ・ジリクナがRoyal Gold Medal For Architecture賞を受賞した。			
4	1980年にガイゼル・コニツがRoyal Gold Medal For Architecture賞を受賞した。			
5	1982年にグルト・シュリョーグルがオーストリア賞を受賞した。			
6	1976年にアルベルト・ツヴァイミュラーがドイツ建築賞を受賞した。			
7	1983年にフランク・Q・ゲーリーがアーノルド・W・ブルンナー賞を受賞した。			
8	1983年にハンス・ホラインがドイツ建築賞を受賞した。			
9	1976年にビセンテ・ミランダがFAD賞を受賞した。			
10	1979年にオスカル・トウスケがFAD賞を受賞した。			
11	1976年にフェルナンド・アマトがFAD賞を受賞した。			
12	1983年にノーマン・フォスターがRoyal Gold Medal For Architecture賞を受賞した。			
13	1984年に杉本貴司が毎日デザイン賞を受賞した。			
14	1985年に黒川雅之が毎日デザイン賞を受賞した。			
15	1987年に内田篤之が毎日デザイン賞を受賞した。			
16	1972年に倉俣史朗が毎日デザイン賞を受賞した。			
17	1984年にフアン・ウ・アンド・ハートマン・アキテクトがAIA賞を受賞した。			
18	1981年にH.E.D.アキテクトがAIA賞を受賞した。			
19	1980年にマイケル・ロトンディがAIA賞を受賞した。			

図7 連想的ブラウザ