

# テキスト情報の可視化を 利用した情報検索

情報の可視化あるいは視覚化 (information visualization) と呼ばれる技術は、1990年代以降に独立した研究分野として広く認識されるようになり、ここ数年で急速な発展をとげたものである。CADや科学技術計算の可視化によって成熟した技術が、安価で高性能なハードウェアを利用して大量のビジネス・データやリソースを効果的かつ効率的に検索・把握し、処理することを可能にした。

本稿では、このような情報の可視化に関する概念と技術を紹介し、特に大きな成功を収めている情報検索への応用について詳しく説明する。

日本アイ・ビー・エム (株) 東京基礎研究所  
武田 浩一 ・ 野美山 浩

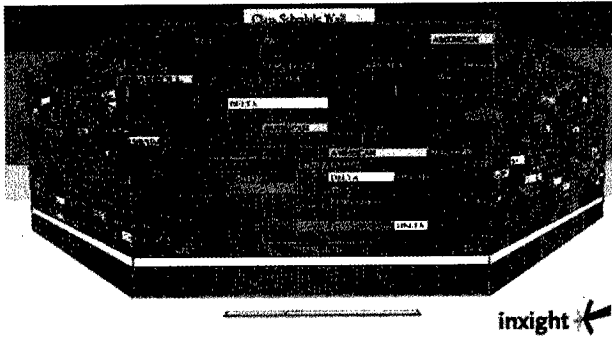
## ■ 情報の可視化 —

可視化とは本来直接目で見るできない情報を何らかの形で見えるようにすることである。従来は科学技術計算の分野で、シミュレーション結果あるいは属性値の範囲や分布を理解しやすい形で表現する技術 (scientific visualization) を中心に研究されてきた。このような可視化の特性は、同じ情報でもその提示やデータ表現の選択によって我々の認知容易度が異なるということであり、たとえば Bertin<sup>1)</sup> や Tufte<sup>12)</sup> によって早くから指摘されてきた。したがってさまざまな種類の情報やタスクに応じて、最適な情報の可視化手法を選択できるように豊富な知見を蓄えることが重視されるようになった。たとえば、ユーザが記憶あるいは処理可能な資源の拡大、検索作業の負担の軽減、特定の情報パターンの検出支援、注意喚起によるモニタ支援、情報の視覚的な直接操作による使いやすいインタフェースの提供、といった目的で可視化手法が考えられてきた。

1989年になって Xerox PARC (Palo Alto Research Center) の Robertson らは、プログラム、文書ファイルやディレクトリといったオフィスのリソースを対象とした情報空間 (information workspace) を管理するプロトタイプとして Information Visualizer<sup>3)</sup> を発表した。

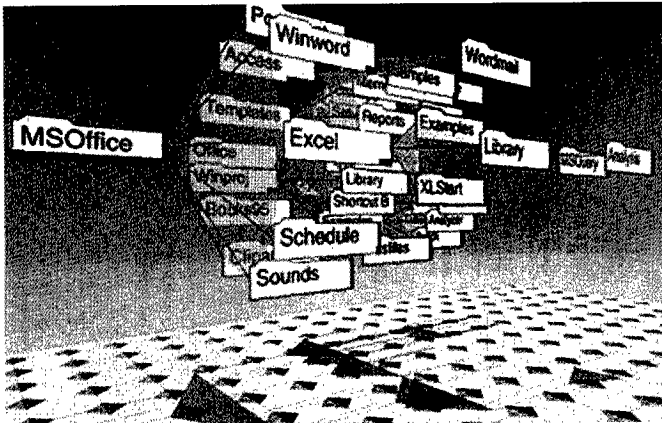
Information Visualizer は、プログラムやファイルのタイプ別の更新履歴の時系列を壁のように表現したパースペクティブ・ウォール (perspective wall, 図-1) や、ディレクトリの階層構造を回転する3次元の木のように表現した円錐木 (cone tree, 図-2) といった多様な可視化表現を採用し、1990年代の情報の可視化技術の先駆的な研究となった。その後は時系列、階層構造、リンク構造といった基本的なデータ構造や、大量のデータ集合を概観 (overview) したり、対話的に探索 (ナビゲート, navigate) するためのさまざまな可視化手法が発表され現在に至っている。

情報の可視化の特徴は、このように抽象的な実体の集合を対象としていることであり、必ずしも数値的属性や地理的属性を持たない情報を効果的に可視化することで広汎な分野への適用が可能となった。たとえば電子図書館における文書検索や著作の支援機能<sup>9)</sup> や、プログラム開発<sup>5)</sup> のためのユーザ・インタフェースとして情報の可視化が利用されている。また可視化とは情報の静的なスナップショットではなく、インタラクションによってさまざまな側面や抽象度から表現される情報が動的に変化することを前提としている。特に情報検索のように必ずしもユーザが事前に検索対象についての明確な知識や定義を持たない場合には、システムからのフィードバックにより段階的に検索要求を改良することがきわめて



縦軸に所有者，横軸に作成日に対応させ，ファイル・タイプを色別にして3種類の属性を表現している。注目している期間が前面になるように表示される。

図-1 パースペクティブ・ウォールによるファイル履歴の可視化  
(courtesy of Xerox PARC)



左端にルートがあり，右に向かってディレクトリとそこに含まれるファイルを展開している。3次元空間の奥行きを利用して，ファイルが平面的に広がらず，かつ注目している部分が前面になるように表示される。

図-2 円錐木によるディレクトリ階層の可視化  
(courtesy of Xerox PARC and Inxight Software)

重要である。可視化はこの点で理想的なフィードバック情報提供の手段となっている。技術的には

- ・可視化の対象となる情報のモデル化
- ・可視化に用いる視覚的なデータ表現の設計
- ・モデル化された情報から視覚的データへのマッピング
- ・視覚的データに対する操作

といった手法・アルゴリズムとそのアプリケーションが主要な研究テーマである。

情報の可視化は主としてHCI (Human Computer Interaction) 分野で注目を集めていたが，研究の高まりを反映してIEEEは1995年にSymposium on Information Visualization (InfoVis) を，また1997年にはInformation Visualization 国際会議 (IV) を発足させ，情報の可視化は独立した研究分野としてより活発な研究

報告が行われるようになった。1998年には，メリーランド大学のShneidermanが彼の著名な教科書「ユーザ・インタフェースの設計」第3版<sup>10)</sup>を出版し，第15章すべてを情報の検索と可視化の解説にあてている。さらに1999年になって，Xerox PARCのCardらの編集による主要論文集「Readings in Information Visualization」<sup>3)</sup>が刊行され初学者にとっても非常に学習が容易になっている。同文献では主要な論文が，可視化される情報空間，対話的処理，文書の可視化，可視化に基づく作業空間やツール，思考過程の視覚的な支援などに分類されており，情報の可視化が，情報検索や認知科学といった複数の領域にまたがる学際的な研究分野であることをよく反映している。

## ■ 情報検索への応用

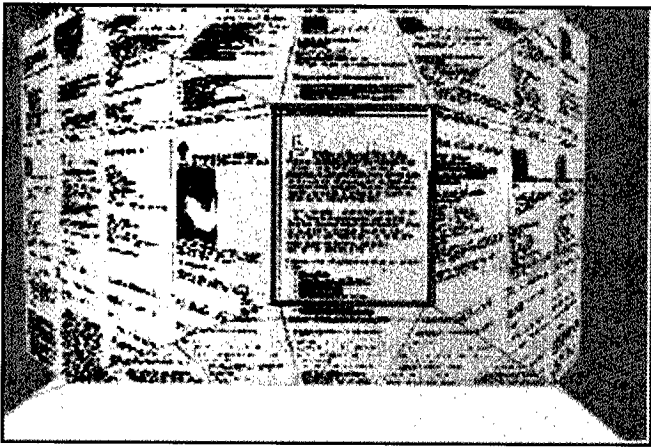
Shneidermanによれば，情報検索は要素の母集団から求めるものを選択するタスクであり，このための情報の可視化手法は2つの分類の組合せ「データ型×タスク」によって特徴づけることができる。データ型には，

- ・線形 (1-D Linear) : テキストやソースコードといった1次元の要素の並び
- ・2次元 (2-D Map) : 地図のような領域を持つ要素の集まり
- ・3次元 (3-D World) : 実世界の実体のように3次元の広がりを持つ要素の集まり
- ・時間的 (Temporal) : 線形の要素の中で特に時系列に関するもの
- ・多次元 (Multi-dimensional) : n個の属性を持つ関係データベースの組のような要素の集合
- ・木 (Tree) : 根となる要素以外は，各要素がただ1つの親要素を持つような集合
- ・ネットワーク (Network) : 木よりも一般的な，要素間の任意の接続関係を持つ集合

の7つがあり，タスクにも

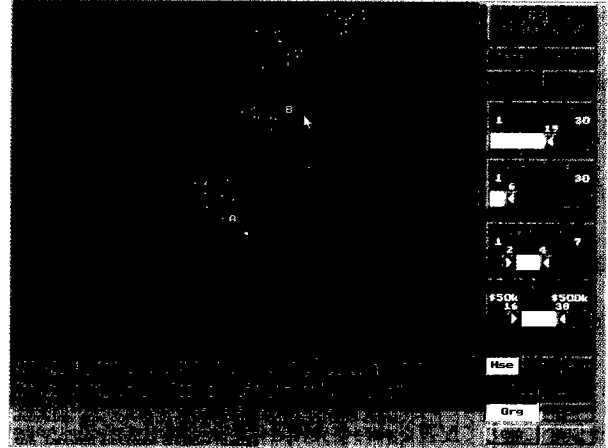
- ・概観 (Overview) : データ集合の全体像を把握する
- ・ズーム (Zoom) : 注目すべき要素や部分を拡大する
- ・フィルタ (Filter) : 興味のない要素や部分を除去する
- ・オンデマンド詳細化 (Details-on-demand) : 個別の要素や特定のグループから詳細な情報を得る
- ・関連づけ (Relate) : 要素間の関係を表示する
- ・履歴 (History) : ユーザの操作履歴を管理する
- ・抽出 (Extract) : セッション管理や検索結果の保存を行う

の7つがあげられている。今までに知られている情報の可視化手法やシステムは，このような観点から大まかに整理することができ，さらに個別のより詳細な機能や特徴による細分化が可能である。たとえば，図-1のパース



中央のページにレンズが置かれている。周辺のページは焦点から外れるにしたがって歪められ (distortion), 内容ではなくページのレイアウトに関する視覚的な文脈情報を提供している。

図-3 文書レンズ  
(courtesy of Xerox PARC)



ワシントンDC地区で、ある2つの地点 (A, B) から特定の距離内、寝室数、価格の3つを条件とした売家の動的質問の実行情例。

図-4 動的質問

ペクティブ・ウォールは、時間的データの概観を、図-2の円錐木は木型データの概観を行うための可視化手法として大別できる。

情報検索において可視化が成功するパターンとして、Shneidermanはさらに

まず概観し、ズームやフィルタの後、オンデマンド詳細化を行う

というビジュアルな情報検索のマントラ (visual-information-seeking mantra) を提唱している。カリフォルニア大学バークレー校のHearstは、情報検索の分野における可視化の重要性を強調し、ユーザが検索途中で得られるフィードバックによって興味の対象や検索目的を変化させる (berry-picking と呼ばれる) 行動、検索結果の利用目的の多様性 (傾向分析や比較など)、ユーザの検索式作成の習熟度の違い、といった情報検索特有の問題を指摘したうえで、さまざまな可視化手法を評価している<sup>4)</sup>。

以下では、これまでに研究開発された典型的な情報検索の可視化システムをもとに、その有効性を検証してみたい (以下で参照したシステムやその手法に関する論文は、文献6) に収録されている)。

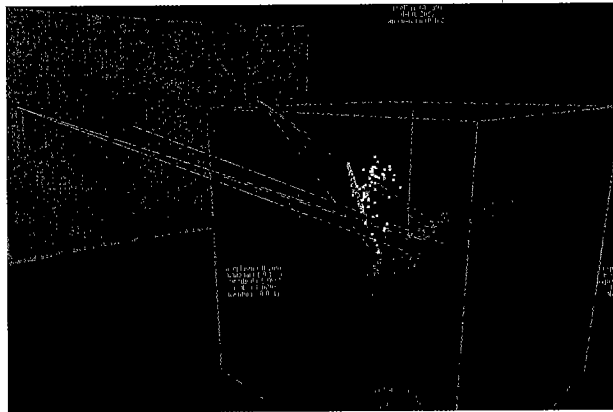
## ■ 文書レンズ

文書レンズ (document lens, 図-3) は文書をページ単位でコンピュータ上で概観するために、2次元平面上に配置したページを3次元的な遠近法で投影するものであり、ユーザの興味あるページを読むための道具としてレンズが用意されている。ユーザは各ページとそれに隣接するページを俯瞰することで、図が多い、表と

その説明がある、といった大域的な特徴をもとに興味のある個所を特定することができる。あるいはキーワード検索の結果、指定したキーワードがどのようなページに分布しているのかを確認することも可能である。レンズを置いたページ (focus) は画面の中央におかれ、文字が読めるサイズに拡大 (ズーム) される。また、魚眼レンズ (fisheye) のように中央ページの周辺は歪められ、注目しているページを邪魔しない程度の文脈 (context) を提供する。このような手法を focus+context 手法と呼び、ブラウジングやナビゲーションのための大域的な情報を保持しつつ、興味のある情報を必要な詳細度で参照するために有効であることが知られており、リンク構造や木構造の可視化などにも応用されている。本手法は、Information Visualizerの拡張として研究開発されたものであり、その後さらに同じサイトの関連Webページの可視化を行う WebBook というシステムにも採用されている。Bierらは、ユーザがレンズをより柔軟に制御でき、またレンズによって注目している部分にさまざまな情報を表示できるようにした magic lens というツールを開発している<sup>2)</sup>。

## ■ 動的質問

動的質問 (dynamic query, 図-4) は、Shneidermanによって提案された手法で、スライダーを利用してユーザがデータベースの検索条件を連続的に変更し、その結果起こる検索結果の画面上の表示の動的な変化をもとに最適な解の集合に到達するという仕組みになっている。それまでの比較的データ集合の静的な観点の可視化に比べて、動的質問では、アニメーション効果とでも呼



類似度に応じて配置された中央付近のデータ集合の各要素が地図上の特定の地点に関連づけられている。

図-5 Starlightの作業空間

ぶべき検索結果の連続的な変化を可視化の観点として、ここに特色がある。ここでは、概観を制御するとともに不必要な情報を除去するフィルタの役割をスライドバーという可視化要素によって実現させている。また地図などを背景として、検索結果を輝点で表示すると視認性もよく色別の情報分類も可能となり、有効な可視化を実現できる。

このような動的質問の可視化手法は、OLAP (On-Line Analytical Processing) と呼ばれるデータベースやデータウェアハウスの分析手法と結びついてより強力な対話的データ分析機能を実現するものと考えられる。大規模なデータベースに対して高速に動的質問を処理することは困難であるかもしれないが、検索結果がある程度絞り込まれている場合には有効な手法といえる。このことは、対象データのサイズや可視化手法のスクラビリティに応じて、複数の可視化戦略を組み合わせる最適なインタフェースを構築するという要求につながっている。テキスト情報に対する動的質問の例としては、映画のタイトルや俳優などで検索を行う FilmFinder というシステムが試作されている。

## ■ Starlight

Starlight (図-5) は、テキストやデータベースを含むマルチメディア情報源を利用した課報活動 (intelligence) を支援するために、さまざまな観点から情報を関連づけ、専門的な分析ができるように米国パシフィック・ノースウェスト国立研究所で開発されたシステムである。複数の可視化手法をワークスペースと呼ぶ3次元の仮想空間上に展開し、それぞれの可視化表現にまたがる要素間の関連はリンクとして明示することができる (このような手法は brushing and linking と呼ば

れることがある)。リンクには文書間の類似度のように、動的に計算されるものや、URL のような直接的参照によるものが含まれる。階層的なデータモデルとリンクによって、ある要素に関連する要素をディレクトリのように入れ子状に展開し、可視化することもできる。これを関連検索 (association query) と呼んでいる。

## ■ 自己組織化マップ

自己組織化マップ (self-organizing map) とは、もともと Kohonen が N 次元の属性を持つ要素を2次元の格子状の平面に配置するために提案した学習アルゴリズムによる情報の可視化であった。Lin はこれを類似文書の分類 (クラスタリング) に利用し、文書集合を特徴的なキーワードとともに2次元平面のそれぞれの領域に分割・配置することに成功した。Lin はさらに、このようにして得られた文書のマップに、たとえば個人の所有する文書の情報を重ね合わせることで、その個人の研究分野や興味範囲を効果的に可視化できることを明らかにした。Kohonen も WEBSOM と呼ばれるインターネット上の文書を自己組織化マップにより可視化するシステムを開発している (<http://websom.hut.fi/websom/>)。図-6 では、WEBSOM によってインターネットの sci.lang というニュース・グループの 32,627 記事を可視化した例を示している。

このようにそれぞれの可視化手法において、注目すべき観点からの概観による全体像の把握と、対話的なズーム・フィルタ機能が巧みに組み合わせられており、さらに完成度の高いシステムや実用的なプロトタイプでは、個別の要素に関するオンデマンド詳細化の機能が備えられていることが分かる。

情報検索における概観のための可視化では、クラスタリングの可視化手法を用いるものが多い。また従来からよく使われる書誌情報や最近の電子図書館などで活発に議論されているメタデータ (データに関するデータ) などの種々の属性情報を利用することにより、文書間の参照関係、著者、タイトル、価格といった豊富な情報に基づく可視化が可能になってきている。システムが自動的に類似文書を判定するクラスタリングでは、事前に分類すべきカテゴリやその文書例が与えられているカテゴリーライゼーションの手法とは異なり、それぞれの類似文書がどのような話題に関するものか、また文書全体がどのような体系で分類されているかといった情報も生成しなければならない。したがって、クラスタリングの可視化で重要なことは、各クラスタに含まれる文書の特徴づける単語やフレーズを自動的にかつ的確に生成する技術である。これはテキストや文書集合を要約する技術という大きな枠組みでとらえることができる。メタデータは、従

来はコンテンツに対して人手で付与するのが一般的であったが、テキストから自動的に人名や会社名といった意味あるまとまり (named entity) を抽出する技術によって補完することができる。最近の研究には、人手でテキスト自体に有用な情報を注釈 (annotation) として付与することで、従来よりもかなり高い精度で要約や翻訳を行うというアプローチがある。情報検索や可視化についても、このような手法が有効であることは疑い得ない。

従来の情報検索がユーザからのフィードバックによるボトムアップ的な検索条件の改良により検索精度の向上を図っているのに対し、可視化に基づく情報検索では概観からトップダウン的に興味のある文書集合に到達する手法が中心となる。可視化によって従来の検索式では表現しにくい検索条件をユーザが容易に指定できるケースも増えてきている。ただし、今後はHearstも指摘するように、ユーザから関連文書あるいは追加検索語のフィードバックが容易になるような可視化を中心に、よりズームおよびフィルタ機能が充実するように技術的要請が高まると予想される。また、最近の研究動向では、レンダリングなどのいわゆるグラフィクス分野の技術も情報の可視化手法と相互に影響しあっていることが分かる。たとえば、文書の内容をそこに含まれる単語に基づいてウェブレット変換し、波形の可視化によって関連した話題や新たな話題の展開などを表現する手法が知られている。テキスト情報の可視化も、単純にクラスタを特徴づける単語や類似度に基づく単語のネットワークを表示するだけでなく、主語や述語からなる文単位の情報を巨大なグラフとして可視化するアプローチが現れている。この手法が成功すれば、自然言語による質問に対して、膨大なテキスト情報から推論した結果を可視化するような高度な応用が可能になる。

情報の可視化は国内の研究者によっても盛んに研究されている。特にソニーコンピュータサイエンス研究所の増井 (LensBarと呼ばれる focus+context手法など) や暦本 (Information Cubeによる3D可視化手法) には可視化手法に関する多数の研究報告がある。増井の情報視覚化の最近の研究動向についての解説論文<sup>6)</sup> は、日本語で書かれたこの分野の貴重な文献となっている。電気通信大学の小池研究室ではFractal Viewsという focus+context手法とソフトウェアの可視化を研究している。他にも神戸大学田中研究室 (データベースの可視化)、富士ゼロックスの植田らによるCastingNet (ハイパーメディアの可視化システム)、東京大学の館村研究室や日立の丹羽による文書中のキーワードと文書空間の可視化、慶應義塾大学の塩沢によるWWW空間の可視化などの研究例がある。著者らも以下で述べるイ

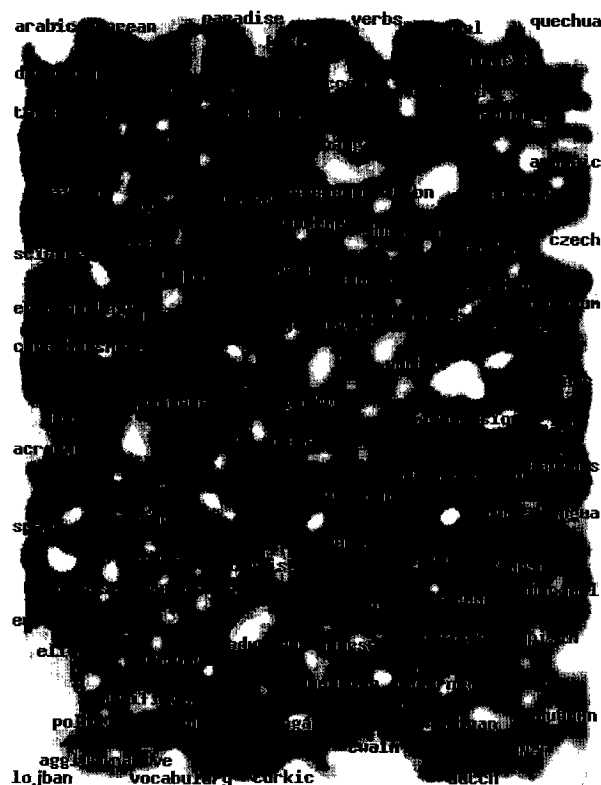


図6 自己組織化マップ  
 文書の特徴をキーワード集合によって表現し、学習アルゴリズムによって類似文書を近傍に配置する。類似文書の集まりであるクラスタを特徴づける単語を自動的に選択・表示し、文書の密度をクラスタの色の濃淡に対応させている。

図6 自己組織化マップ

ンフォメーション・アウトライニング<sup>7)</sup> という可視化手法を1994年から研究開発している。

### ■ インフォメーション・アウトライニング

文書集合を概観するための可視化では、全体像を把握できる程度の抽象度の高い情報を提示する必要がある。しかし特定の可視化では表現される情報に限りがあり、元々の文書集合が持つ豊富な特性をとらえきれないことが多い。たとえば、1999年の1年間の新聞記事の集合から「2000年問題」に関する記事を検索した場合に、月別の該当記事の件数を見れば、それが年末に向かって次第に増加していることが分かる。すなわち2000年問題に関する記事は人々の注目度が高まっている、という重要な情報を得ることができ、2000年になってからは記事数が減少し、話題性が急速に失われているといったことも理解できる。こういった記事数と話題性の関係は、大まかな記事数変化の特徴パターンとして記述することができる。さらに月別に記事によく現れるキーワードの分布や、ある月に集中して現れるキーワードを表示すれば、話題変化や2000年問題のみに固有の話題について



1994年のインターネットに関する158件の記事を検索し、記事に含まれる都道府県名の出現頻度と記事の月別件数分布を表示している。都道府県別では、東京都や兵庫県が多いことが分かり、月別分布からは、インターネット関連記事が1994年に大きな話題になっていった様子が分かる。

図7 インフォメーション・アウトライニング

知ることできる。このようにして関連度順に並べられた記事のタイトル一覧や、単にキーワード別に分類された記事のディレクトリ表示などとは異なる観点からの概観が可能となる。極端な表現をすれば、文書集合に付与されたメタデータと、検索式を満足する文書集合から動的に計算される上記のような特徴だけを利用して、本文を一切読まないでも望みの文書集合に絞り込めるような可視化が目標となる。著者らはこのような豊富な概観（ビューと呼ぶ）をもとに、情報のズーム／フィルタ機能と、ナビゲーションを支援する機能とをビュー上の直接操作として実現し、これをインフォメーション・アウトライニング（情報の輪郭づけ）と呼んだ。直観的にはプログラミング言語smalltalkで採用されたモデル・ビュー・コントローラの3層構造を情報検索向けにアレンジした概念とすることができる。

このように同じデータ集合を異なる観点から可視化するというアイデアはStarlightにも共通するものであるが、インフォメーション・アウトライニングでは

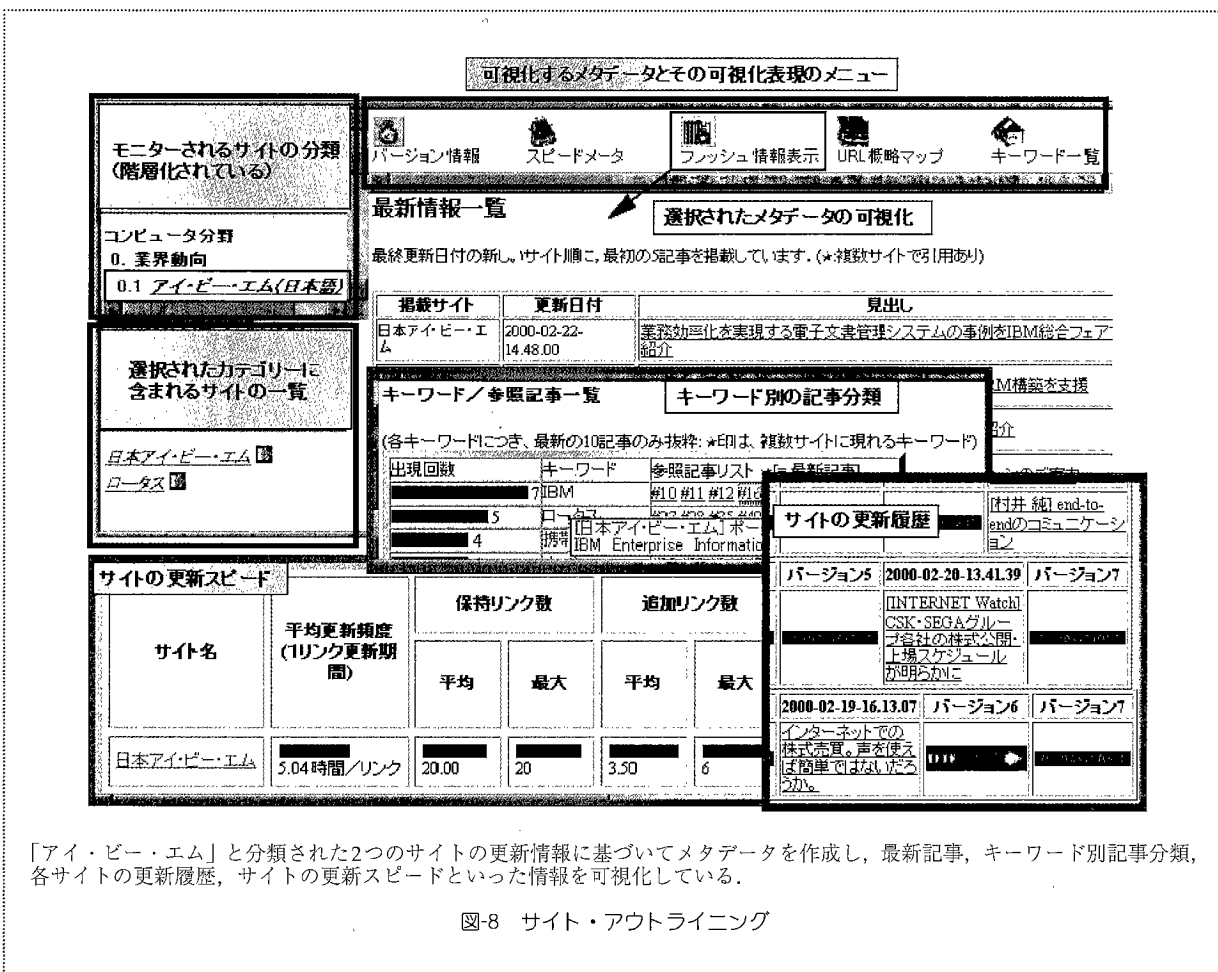
- 要素間の個々の関連よりも、現在注目している要素集合全体あるいはそれらの部分集合の和や差がどういった特徴を持つかを多面的に可視化する
- 動的質問と同様にある観点に対応するビューでの直接操作によって他の観点のビューがどう変化するかを

効果的に可視化する

- 互いに関係するビューを関連づけ、特定の分野やタスクごとに異なるビューの集合を設計してソリューションを実現する

ことを重視している。このようにして、対象となるデータ集合とそのメタデータに応じて柔軟に可視化表現をビューとして選択することができる。ビューは、それ自体が独立した実体として互いに関連づけられたり、可視化できるデータの一般性や詳細度によって上位・下位関係が定義できる。複雑なタスクも比較的単純なビューの集合を組み合わせることで、ユーザにとって使いやすいシステムを構築することができる。インフォメーション・アウトライニングは実際に行政情報、新聞記事、テレビニュース、Webページなどのさまざまな情報に対して可視化に基づく大規模な情報検索ソリューションを構築する手法として採用されている。図7は、インフォメーション・アウトライニングのプロトタイプシステムによって1994年の1年間の日経新聞記事約18万件からインターネットに関する158件の記事を検索し、検索結果に含まれる都道府県名の出現頻度（左下のビュー）と月別件数分布（右下のビュー）を可視化したものである。画面上部中央には記事のタイトルが表示されている。

ビューの多様性はメタデータの豊富さに大きく依存す



るので、人手で付与される書誌情報以上のメタデータを可視化するためには、テキストからの情報抽出やテキストマイニング<sup>8)</sup>の技術がきわめて重要である。たとえば、世界地図上に各国別の参照記事数やエリアごと(北米、東南アジアなど)の記事数を表現するためには、各エリアに含まれる国名といった語の階層関係の定義(シソーラス)が必要であり、さらにエリアごとの傾向を把握するための統計的処理が必要となる。テキストマイニングでは構文レベルの解析によって「何が+どうした」という係り受け情報を抽出し、問題分析や意思決定のための情報抽出を行っている。これらの技術によって大量の文書集合に対しても実用的な可視化が実現できる。動画像を含むマルチメディア文書のように個々の文書のサイズが大きな場合には、このようなメタデータ(内容の要約、キーワード、更新日、作者、代表的なシーンのサムネイル画像など)を効果的に利用することでスケーラブルかつ多数のクライアントの検索要求に耐える情報検索システムが構築できる。

## ■ インターネット情報の可視化

インターネット情報の可視化という観点からは、通

常の情報検索のための可視化に加えて、

- 参照リンクを持つハイパーテキストとしてのWebページ
- あるURLを起点としたWebページの集合としてのWebサイト

の2種類の可視化手法が特に研究されている。前者の場合には検索条件に合致するWebページの集合を求めただけでなく、検索結果のWebページと参照・被参照の関係にあるWebページの集合を求めたり、多数のWebページと参照・被参照の関係にあるWebページの重要度を上げて順序づける方法が知られている。後者の可視化としては、Webサイト紹介のためのヘルプ機能として提供される、サイトマップのさまざまな手法などが含まれる。このような可視化は、一般にページを節点に、リンクを枝に対応させたグラフ構造の可視化手法と考えられ、focus+context手法を適用したものが多数存在する。たとえば冒頭で紹介したXerox PARCのシステムを商品化しているInxight社では(<http://www.inxight.com>)、サイト・レンズと呼ぶサイトマップ構築ツールを提供している。世界中に数億ページあると予想されるWebページの検索エンジンは、おそらく最も多くのユーザが利用し、かつ最も高い性能が要求される情報検索サービスの1つであると考えられ、残念ながら負荷のか

かる可視化手法はほとんど採用されていない。いわゆるポータル・サイトの記事や情報サービスのレイアウトを個人向けにフィルタあるいはアレンジする手法を採用するものが多い。研究レベルでは、これまでに紹介したようなWebBookやWEBSOMなど多くの可視化が存在する。

Webページのメタデータに関しても、以下のようなインターネット特有のものがある。

- Webページの内容を多くのユーザが評価することで、そのページの有用度が推定できる
- Webページは内容の更新頻度が高いページが多いため、更新履歴や最終更新日時が重要である

前者は、協調フィルタリング (collaborative filtering) と呼ばれる技術で、AT&TのHillらによる、ニュースグループ記事の評価をメタデータとして可視化を適用したシステムPHOAKS (<http://www.phoaks.com>)などで使われている。後者の例としては、いわゆるプッシュ技術によって更新された内容を記事単位でユーザに提供するシステムや、フレッシュアイ (<http://www.fresheye.co.jp>) のような過去1か月以内に更新された情報のみを対象としたインターネット検索エンジンに利用されている。また、インフォメーション・アウトライニングをWebサイトに適用し、最新情報や更新履歴を可視化する (サイト・アウトライニング) 手法も報告されている<sup>11)</sup>。図-8は、サイト・アウトライニングのプロトタイプシステムの画面例を示している。左上にある個人別のディレクトリ形式のサイト階層分類から好みのカテゴリ (ここでは「アイ・ビー・エム」) を選択すると、そのカテゴリに含まれる実際のWebサイトの一覧 (中段左) と、それらのWebサイトの更新情報に基づいて最新情報 (新たに追加されたリンクやテキスト)、一定期間に掲載された情報のキーワード分類、各サイトの更新履歴、各サイトの更新頻度や更新される情報量などを可視化できる。

## ■ 今後の展望

情報検索には特定の記述を検索したり、ある事実の有無のみを確認するような利用方法 (特定検索) と、多くのデータから観察できる特徴や傾向を把握したり、単に最近の情報を知りたいためにブラウズするといった利用方法 (広域検索) がある。可視化手法も特定検索と広域検索という2つの要求に答えるように発展を続けることが期待されている。特定検索では、目的や対象が明確であり、できるだけ効果的な検索式の修正や検索結果の絞込みを支援する必要があるが、後者の場合にはサンプリング、要約やナビゲーションといった機能が重視される。特定検索では、従来から適合率や

再現率といった尺度によって検索技術の評価手法が確立しており、学習用あるいは評価用データも入手できる。広域検索に関しては、こういった評価法やデータの確立が望まれる。

インターネットにおける情報検索と可視化については、さらに独自の進展が期待できる。情報量という点では大差がないWebページでも、デザインやインタフェースのよさによってユーザへのアピールが大きく異なるように、可視化手法もアートの観点や広告手法を考慮したものが注目を集める可能性がある。インターネットにおけるポータルをはじめとした情報検索サービスは、ショッピングや広告といった商業的な側面が非常に重要であり、このような方向で可視化手法を考えることで社会的に大きなインパクトを与えることができるかもしれない。年齢、性別、趣味などのあらゆる面で異なる層のユーザからの情報要求に答えるインタフェースを構築するためには、情報を要求する側だけでなく、提供する側からも豊富なメタデータが必要である。

メタデータの観点からは、さらに口コミに似た主観的なコンテンツ評価データやボランティア活動のような形でのコンテンツ/メタデータ作成が充実することで今まで以上にユーザが恩恵を受けるようになるかもしれない。検索エンジンは客観的なデータだけでなく、さまざまな主観的データを取り入れて可視化とともにマッチ・メイキングのような機能に進化するのかもしれない。感性情報処理との融合も期待できよう。

## 参考文献

- 1) Bertin, J.: *Semiology of Graphics: Diagrams, Networks, Maps* (Bergh, W. J., Trans), University of Wisconsin Press (1967, 1983).
- 2) Bier, E. A. et al.: A Taxonomy of See-Through Tools, Proc. of ACM Conf. on Human Factors in Computing Systems (CHI), pp.358-364 (Apr. 1994).
- 3) Card, S. K., Mackinlay, J. D. and Shneiderman, B.: *Readings in Information Visualization: Using Vision to Think*, Morgan-Kaufmann (1999).
- 4) Hearst, M.: Chapter 10: User Interfaces and Visualization, in *Modern Information Retrieval* (Eds. Baeza-Yates, R. and Ribeiro-Neto, B.), Addison-Wesley, pp.257-323 (1999).
- 5) 小池英樹: ビジュアルライゼーション, bit別冊ビジュアルインタフェース-ポストGUIを目指して- (平川, 安村編), 第2.1章, 共立出版, pp.24-44 (1996).
- 6) 増井俊之: 情報視覚化の最近の研究動向, 情報学シンポジウム招待論文, pp.153-160 (Jan. 2000).
- 7) Morohashi, M. et al.: Information Outlining - Filling the Gap between Visualization and Navigation in Digital Libraries, Proc. of International Symp. on Digital Libraries (ISDL), pp.151-158 (Aug. 1995).
- 8) 那須川哲哉他: テキストマイニング-膨大な文書データの自動分類による知識発見, 情報処理, Vol.40, No.4, pp.358-364 (Apr. 1999).
- 9) Rao, R. et al.: Rich Interaction in the Digital Library, CACM, Vol.38, No.4, pp.29-39 (Apr. 1995).
- 10) Shneiderman, B.: *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 3rd Ed., Addison-Wesley (1998).
- 11) Takeda, K. and Nomiyama, H.: Site Outlining, Proc. of 3rd ACM Conf. on Digital Libraries, pp.309-310 (June 1998).
- 12) Tufte, E. R.: *The Visual Display of Quantitative Information*, Graphics Press (1983).

(平成12年2月7日受付)