

# 技術展望 情報アクセスインタフェースとしての要約・可視化と動向情報

東京大学 加藤 恒昭 NTT 松下 光範

## 1. はじめに

我々を取り巻く情報環境は、その膨大な量と多数のメディアということに加えて、品質と発信者の多様性という特徴を加えつつある。このような状況で利用者が必要とする情報を得るという行為は、すまず対話的かつ探索的なものとなっている。つまり、情報の収集は、自分の関心を正確に述べて、その応答として適合する情報を過不足なく得るという一往復のやりとりではなく、漠然とした関心に基づいて収集された情報の全体像を概観し、それを通じて、どの細部が更なる関心であり詳細な情報を必要とする部分であるかを模索し、情報の絞り込みや関心の具体化を対話的に進めていく過程となっている。

最近のゲーム機業界について興味を持つ場合、その概要、例えば、シェアや販売台数の大きな変化等の特徴的な出来事がどの時期にあったかをまず把握し、その後、それらの時期の状況について述べられている記事を捜して精読する。ある製品の評判について関心を持った場合、その評判情報をマイニングして、どの特徴に関する肯定的あるいは否定的意見が多いかという状況を概観してから、ある特徴についての代表的な否定的意見に目を通していく。そのような行動が典型的な情報収集の流れになっている。

本稿では、この収集された情報の全体像の概観とそれに続く絞り込みや具体化、そして適切な情報へのアクセスを支援する仕組みである情報アクセスインタフェースについて、要約と可視化という観点でまとめる。ここでは、収集された情報の全体像を示すことは、それら情報群の可視化、視覚的要約の生成であり、その視覚的要約は、絞り込みや具体化、そして適切な情報へのアクセスという利用者の要求に回答しうる対話的なインタフェースとなっている

べきだと考える。テキスト情報を中心に議論するが、ここでの議論の多くは、画像情報等を対象とした場合にも共通する。次に情報の概観の中で特に動向情報の把握ということに注目し、動向情報の表現が情報アクセスのインタフェースとして興味深いことを述べる。そして、研究事例の紹介を通じて、動向情報の要約と可視化によって、情報の概観とそれに続く絞り込みや具体化、そして適切な情報へのアクセスが支援されることを示す。

以下、まず、情報の概観と情報へのアクセスを支援する要約と可視化を概観する。続いて、動向情報を位置づけ、動向情報の要約と可視化に関する研究の現状を示す。最後に、動向情報の要約と可視化に関する研究を紹介し、その重要性を主張する。

## 2. 要約と可視化

図1は文献[1]に示されている可視化の参照モデルである。これによれば、可視化の過程は、特徴量を抽出するデータ変換、それに視覚的構造を割り当てる視覚的写像、その構造の見せ方を定めるビュー変換からなる。そしてこの3つの過程に利用者が関与することで利用者の関心に沿い、タスク達成に有意味な可視化が得られることになる。

この参照モデルに従って、一定の関心に従って収集された情報群の全体像あるいは概要を可視化することを考える。概観に加えて、その後に情報の絞り込みや関心の具体化が続くという前提でそれを支援することが目的であるので、個々の情報単位あるいは類似情報をまとめあげたものにそれが何を述べているか等の特徴とし、それらの間の関係や分布に視覚的構造を割り当て、組織化するのが一般的である。ただし、情報群が全体として何を述べているか

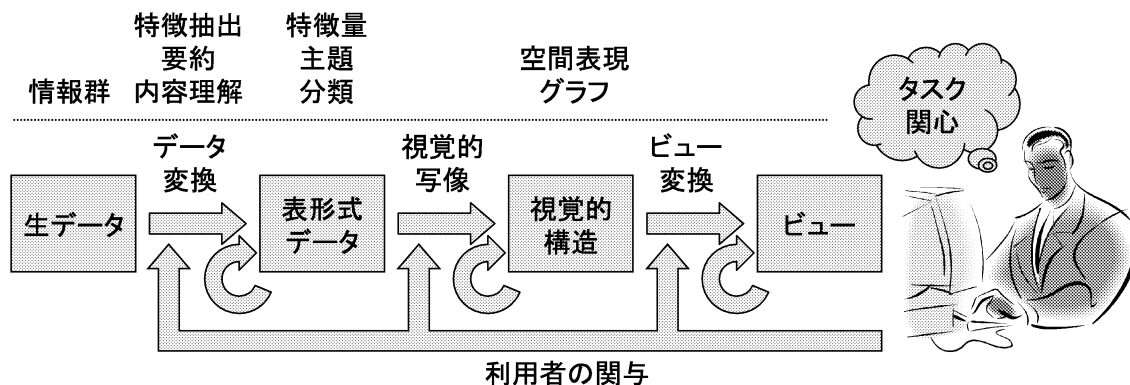


図1 可視化の参照モデル

という要約を視覚的に示すことも可能であり、そのような例は本節の最後で紹介する。

個々の情報からの特徴抽出とその組織化による提示の最も身近な例は、タイトルとキーワード近辺のテキスト断片を情報単位であるページの特徴として、スコアの高い順にリストとして配置する検索エンジンでの検索結果表示であろう。しかし、スコア順に組織化されているとはいえ、情報単位の間の関係は把握しやすいものではない。より高度な組織化として、情報の特徴として主題を抽出し、それに空間という視覚的構造を与え、主題や情報を空間に配置するものがある。様々な手法のヴァリエーションがあるが、情報である文書や段落に語や句が現れた頻度や、同一文書内や同一文内にとどのくらいそれらが同時に現れたかという共起頻度を用いて、ある情報が何に関するものであるかという主題もしくはテーマに沿うように情報群をクラスタリングすると同時に、それらの間の距離を計算し、それに基づいて空間中に配置するのが基本的流れである。情報単位が配置される場合や主題として抽出された語が配置される場合がある。このような空間に分布した情報群あるいは主題群を眺めることで、利用者はどのような主題を持った情報が多いのか、それらの主題は強く関連しているのか等を知ることができ、関心の詳細化とそれに応じた情報アクセスが可能になる。

このような配置では情報の相対的な位置関係だけが意味を持ち、空間の軸が何を意味するかの解釈は利用者に任せられる。これに対して、視覚的構造として意味付けが外部から与えられる空間あるいは軸を与え、そこに情報を配置することも可能である。代表的かつ基本的なものは、時間の軸と地理的空間である。時間軸への配置はその情報がいつ発信（発行）されたかという書誌情報的なものに基づく場合と、その情報がいつに関してのものなのかという内容に基づく場合とが考えられる。前者はむしろ例外的で、地理的空間への配置はその情報がどこについてのものであるかという内容に基づく配置であるのが普通である。応用に依って様々な軸が存在する。評判情報であれば、肯定的な意見であるか否定的な意見であるかという軸に加えて、製品の価格や機能等、どの特徴に関するものかという軸がある。ヘルプデスクに集まるコメントであれば、何に関するものかという軸に加えて、何かをしたいのか何かが起こってしまったのかというような分類があろう。

ここで軸と呼んだものの一部は、離散的であり順序も意味を持たないので、分類の観点ととらえる方が適切な場合もある。その場合、割り当てる視覚的構造は空間である必要はなく、それらの分類に関するヒストグラム等のグラフも考えられる。評判情報を例にとれば、どの特徴については肯定的意見が多くどの特徴には否定的意見が多いのか、ある特徴にはそもそも関心が持たれていないのか等を可視化するグラフとして、情報の全体像を提示することができる。なお、この場合、グラフの対応する領域が情報と対応

づけられる。例えば、ある特徴についての肯定的意見の件数を表現するグラフの部分を指定することで、その分類の情報にアクセスすることができる。

このように情報を空間に配置し、分類に応じてグラフ化することが視覚的写像である。そして、そのための特徴の抽出はデータ変換とされているが、このような情報群の可視化の場合、データ変換は内容の理解であり要約である。これまで示したように様々な理解と要約が必要となる。テキスト要約においては、基になった文書が読むに値するかを判断するための指示的要約とそれを読む代わりとなるような報知的要約が区別される<sup>12)</sup>。例えば、あるテキストが何を主題とするかが分かれば読む読まないの判断ができるので指示的要約としてはそれで充分だが、それだけではその主題についてどう述べているかがわからない。その内容までを明らかにするのが報知的要約である。しかし、ここでの内容の理解と要約においてはその区別は必ずしも明らかではない。その情報の主題は指示的要約であろうが、価格についての否定的意見であることには報知的要約の側面がある。時間軸、地理的空間に配置するために必要な「10月1日に箱根町で震度5の地震があった」という情報は報知的要約と呼んでよい。どのような特徴に基づいて情報群の概要を示したいかに応じて様々な理解と要約が必要となる。加えて、可視化された情報群中における個々の情報への注釈として、特徴抽出とは違った観点での要約が利用されることもある。

このように配置された情報をどのように見せるかがビュー変換であるが、周辺の情報を持ち捨てず注目点を詳細に眺めることを可能とする魚眼レンズを通した見せ方を用いて、時間の軸と主題の軸からなる空間に配置した有名な Perspective Wall<sup>13)</sup>を筆頭に様々な手法がある。物理的な空間を眺めるのと同様にズーミングやアングルの変更によって様々な見方を許したり、時間と主題、時間と地理的空間のように異なる特徴に基づいた複数の空間中での情報群を同時にあるいは切り替えながら眺めることで、情報の新しい特徴を見つけ出すテキストマイニングの支援もこの枠組みの中で可能となる。これらに加えて、表示される情報群そのものを絞り込むような操作や個々の情報を詳細に眺める操作により、収集された情報の全体像の概観とそれに続く絞り込みや具体化、そして適切な情報へのアクセスが進められる。

これらとは異なるアプローチとして、複数文書要約技術によって情報群全体の要約を作成することで、文字通りの概要を示す試みもある<sup>14)</sup>。複数文書要約システムに対話的なインタフェースを持たせるという提案であるが、要約率やどのトピックを含めるかを指定するという操作によって、様々な要約を作成することができ、得られた要約の部分部分それぞれがどの文書から抜き出されたかの情報を持つことで、要約からその基となった文書へのアクセスを支

援している。アクセスされた文書の各部分はその要約における重要さに応じて濃淡がつけられており、文書内の情報の取捨選択の支援となる。視覚的構造とは呼べないが、要約文書という構造の中にそれぞれの文書から抽出された特徴であるところの重要部分を組み込んでいくという枠組みと考えることで可視化の参照モデルと対応づけることができる。なお、このシステムでは、各文書が言及している場所に基づいて文書を配置することも組み合わせられている。

このように考えていくと、個々の情報から特徴の抽出ということである種の内容理解と要約を行い、それらに視覚的構造を与える等して組織化したものが、情報アクセスのインタフェースとして、情報の概観とそれに続く絞り込みや具体化、そして適切な情報へのアクセスを支援することがわかる。組織化それ自体も要約の一種と考えることもできるので、そのようなインタフェース全体が、情報群の要約と言ってもよい。そしてこの要約は、利用者の様々な要求に対話的に応答する仕組みを有しており、多くの場合、視覚的である。

### 3. 動向情報

情報アクセスのインタフェースとなる情報群の要約となるべきものを考えると、それは、「06年からゲーム機業界はどんな感じになったのか」「今年に入って原油とガソリンの価格はどう動いているのだろうか」「去年の台風はひどかったのか」等、利用者の関心に対する最初の回答となるべきものであろうと想像できる。それらはゲーム業界やガソリン業界や台風のいわゆる「動向」である。もう少し具体的にいえば、幾つかの統計量の時系列データとそれに関する評価や判断を基として、その変化を通時的にとらえつつ、単に羅列するのではなく、総合的にまとめ上げることで得られる要約であるような動向情報である。このような動向情報は、利用者の関心に対する最初の回答であり、それを足掛かりに関心の詳細化や必要な情報の具体化が行われていくという点で、情報の全体像の概観とそれに続く絞り込みや具体化、そして適切な情報へのアクセスを支援する仕組みの核となりうる。加えて、ここであげたような例は、要約と可視化という点で以下の興味深い性質を有している。

- ・一定期間にわたる情報を総合的にまとめあげることが必要であり、情報の間に重複も多く、組織化が重要である。
- ・生産台数やシェアのように時系列データである統計量への言及を含むものや、台風や地震のように地理的な情報を含むものが多いことに加えて、シェアにおける企業、支持率における政党等、情報を配置する軸に様々な可能性がある。
- ・統計量への言及等の事実情報はその一部に過ぎず、そ

れに加えて、その解釈、原因の推測や波及効果の予測等が重要な意味を持つ。

- ・新聞記事のようなテキスト情報に止まらず、統計量に関する詳細な数値情報が白書等に存在し、それらを横断した情報の形式にとられない情報収集が必要となる。

このような動向情報の要約と可視化の重要性に着目し、そしてそれに関する研究が、テキスト要約や情報抽出を中心とする自然言語処理と情報可視化という従来あまり交流のなかった研究領域にまたがっていることを問題視して、筆者らと国立情報学研究所の神門典子氏がオーガナイザをつとめて、「MuST: 動向情報の要約と可視化に関するワークショップ」(以下、MuST)を2005年より開催している<sup>6)</sup>。このワークショップは、動向情報の要約と可視化に関する技術について、協調的かつ競争的に研究を進めていこうというもので、議論と研究の活性化、研究コミュニティの形成、ツールやコーパス類の蓄積を目的とし、これらの目的のために、共通の研究資源であるデータセットを用いて、緩い意味で共通の課題に取り組んでいる。

このMuSTの中では、主に以下のような研究に関心が持たれている。

- ・新聞記事テキストから統計量に関する記述等を抽出する技術：情報の特徴としてそこで言及している統計量を抽出する。この統計量の変化をグラフ化し、それぞれの点がどの情報から抽出されたかを示すことで、統計量のグラフを介した情報群の概観が可能になる。
- ・新聞記事中に含まれる情報を適切に可視化する技術：地震の発生や地価の変動に関して、地理的空間と時間軸に情報を配置する、動向の参考になる時系列情報をその性質に従って適切にグラフ化する、加えて、このような可視化にどのようにテキスト情報の要約を重ねるかが考察されている。
- ・統計量の値の変化、数値情報の列を説明する言語表現を生成する技術：数値情報を文章で要約することでその全体像の把握を支援する。そのような要約で言及されている変化や時点は、情報群の中で重要なものであるため、それらの情報へのアクセスを動機づけることになる。
- ・大量情報の概要把握としてのトレンドマイニング技術：一種のテキストマイニングで、統計量に関するキーワードの出現傾向等を基に、最近の話題であるようなトレンドを抽出して可視化したり、統計量間の関係や、それらがトレンドを作り出しているか等を抽出する。

動向情報の要約と可視化に関するこれらの研究は、すべ

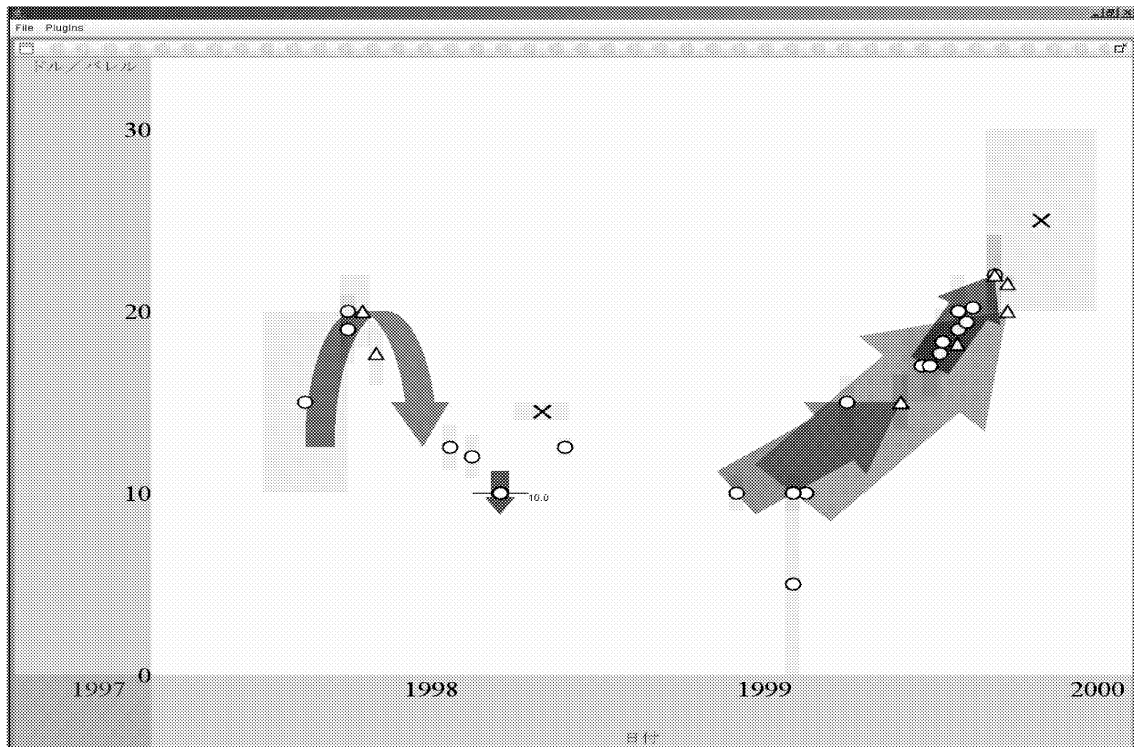


図2 変化の抽出に基づく動向情報の可視化

てが情報アクセスインタフェースの構築を主たる目的としているわけではないが、次節に示すように、これらはそこに多くの示唆を与える。

#### 4. 動向情報の要約と可視化：事例紹介

MuSTの枠組みの中で行われた研究をふたつ紹介することで、動向情報の可視化と要約が情報アクセスのインタフェースとして有意義であることを示したい。第一は、図2に示すような統計量の抽出とそれに基づいた情報アクセスインタフェースである<sup>16)</sup>。このグラフにプロットされている点は、記事テキストから情報抽出技術によって抽出されたもので、それぞれの点がそれを抽出したテキストの部分や記事へと関連づけられている。

特徴的であるのは、様々な形状の矢印記号がグラフに貼付けられている点で、これらは、テキスト中の変化に関する言及、「下落」「低迷」「急騰」等に対応している。新聞記事テキストを見ると、例えば、「原油価格（ドバイ原油）も、昨年10月ごろ1バレル＝約20ドルをつけたのをピークに下落が続き、今年1月下旬に同約12ドル50セントまで落ち込んだ。その後、イラク情勢の緊迫化で一時上昇したものの、現在はまた12ドル前後で低迷、…（毎日1998/2/14）」のように、ある時間幅や特徴的な時点の変化についての定性的な情報が多く含まれており、言語表現としての情報の分節という観点からも、「つけた」「下落が続き」「落ち込んだ」「上昇した」「急伸した」等の変化に基づいていることがわかる。加えて、「昨年10月ごろから下落が続き、今年1月下旬に同約12ドル50セントまで落ち込んだ」と

いような変化に関する情報は、それ自体が3ヶ月間の原油価格の変化についての要約であり、個々の統計量よりも抽象度が高く、統計量の変化の概要、動向を把握するために重要であると考えられる。

変化に関するこれらの言及を含めて、特定の統計量に関する記述を情報、この場合は新聞記事テキストの部分部分、の特徴として抽出し、視覚的構造として時系列グラフを割り当てることで、視覚的な要約を構成している。これによって統計量の変化を中心とした動向の概観を可能とし、利用者は興味深い変化がどの時点で生じたかを知るとともに、それらへの言及にアクセスすることができるようになる。ある時点の統計量を述べている記事だけでなく、ある時期の「低迷」や過去の特徴的な「ピーク」に言及している部分へと直接アクセスできる点がこのインタフェースの利点である。

もうひとつの研究はより汎用なインタフェース構築のツールに関するものである<sup>16)</sup>。図3にそのスナップショットを示す。画面の右側に時系列情報のグラフが描かれており、その幾つかの点にアイコンが配置されている。それぞれのアイコンは画面左側に並べられた情報表示ボックスと関連づけられている。画面の右側のグラフはズームングやパニングの変更が可能で、グラフ全体を眺めたり、その細部に注目したりすることができる。グラフ中に配置されたアイコンを指定すると、必要に応じて画面左側が上下し、対応する情報表示ボックスが画面の中程になりハイライトする。逆にある情報表示ボックスを指定すると、グラフが移動し、対応するアイコン

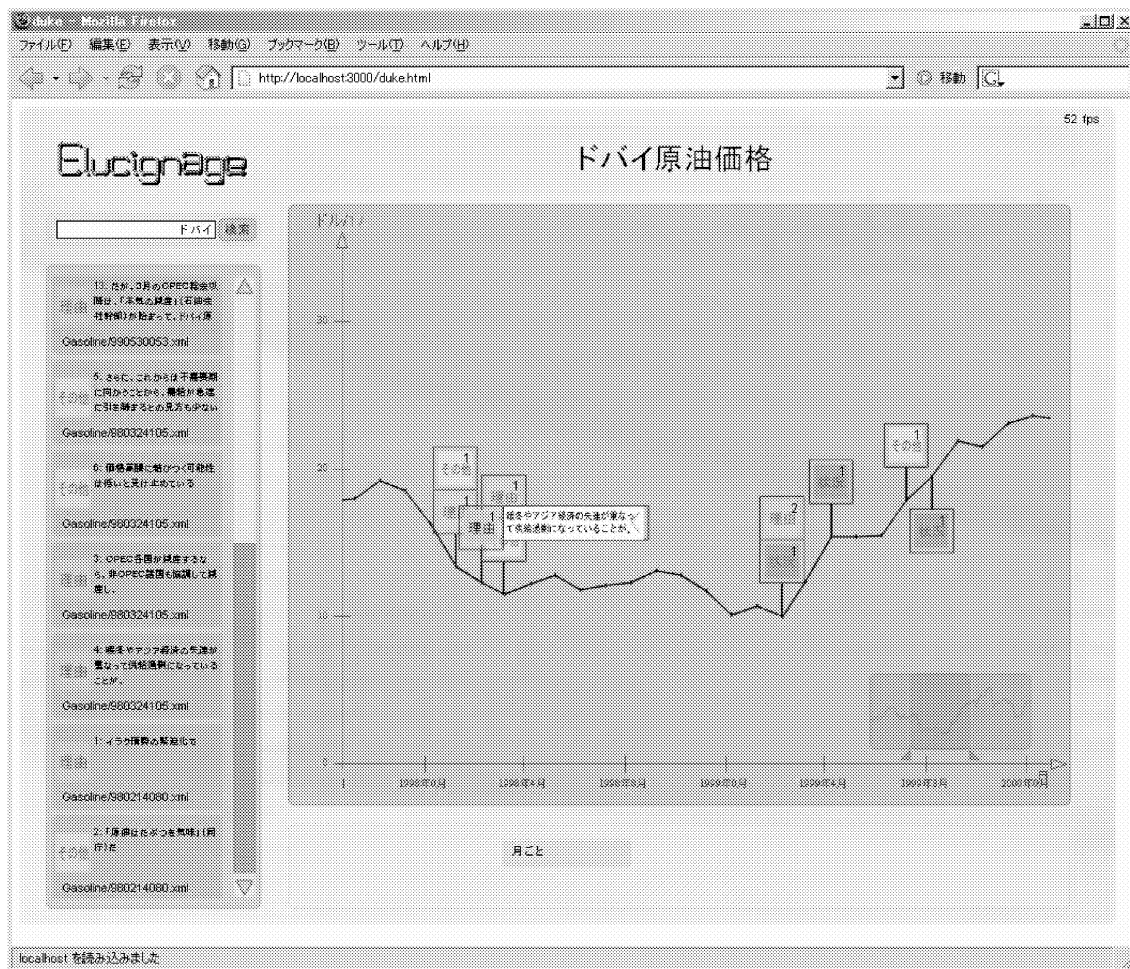


図3 動向情報を中心とした情報アクセスインターフェース

が画面右側中央に位置し、ハイライトされる。

このツールによって実際に構築されるインターフェースの一例は次のようなものとなる。ある統計量に着目し、それに関連する新聞記事等のテキスト情報にアクセスすることを考える。グラフは白書等にある数値情報をデータとして描画する。新聞記事等の情報について、それが描画されたグラフのどの時点に言及しているかに基づいてアイコンを配置する。右側の情報表示ボックスは記事の発行順に並べ、そこに記事テキストを記述する。このような構成によって、先と同様な情報アクセスを可能とするインターフェースが構成できる。利用者はグラフの概観や細部の観察を通じて興味を持ったテキスト情報のアイコンを指定することで、その詳細情報を得ることができるし、逆に発行順に記事を通読しながら、必要に応じてその記事が言及している状況をグラフとして確認することができる。なお、関連する記事や統計情報等を検索してくるモジュールや、記事を適切にグラフ中に関連づける特徴抽出のモジュールは、このツールの外側である。

このツールは画面右側のグラフ、情報表示ボックスの内容や順序に何の仮定もおいてない。グラフに関する基本的な操作と、グラフ中のアイコンと情報表示ボックスとの対

応付けのみをサポートする。であるから、このグラフは白書の数値情報に基づく必要はなく、テキスト情報から抽出された情報に従って描画されてもよい。実はグラフである必要もなく、何らかの特徴に基づいて情報を配置した2次元平面であっても構わない。配置するアイコンも複数の種類が利用でき、そこに意味を持たせることも可能である。加えて、このツールを基に様々なインタラクションを追加することが可能な枠組みになっている。動向情報を中心としつつ、要約と可視化の枠組みに基づく情報アクセスインターフェースの基盤部分をツール化したものといえる。

## 5. おわりに

情報アクセスにおいて必要となる情報群全体の概観とそこから詳細化や具体化、そして個々の情報へのアクセスを支援するものとして、要約と可視化というキーワードでインターフェースについて述べた。動向情報がそのようなインターフェースを構成する情報として興味深いことを示し、その要約と可視化に関する研究を紹介することで、動向情報を中心とする、要約と可視化に基づくインターフェースの有用性を示した。可視化に基づく情報アクセスの支援は新しい概念ではないが、そこに様々な情報の理解と要約とい

う考えを持ち込み、例えば動向の表現を視覚的インタフェースと考えることで、新しい枠組みが見えてくると考える。その可能性に期待したい。

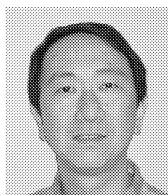
### 謝辞

MuSTのオーガナイザとして筆者らと共にご尽力いただいております。本原稿についても貴重なコメントを頂いた国立情報学研究所の神門典子氏に感謝します。MuST参加者の皆様には、様々な機会を通じて、いつも貴重な示唆を頂いています。お礼申し上げます。

### 参考文献

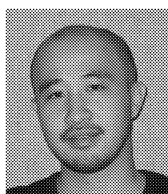
- [1] Card, S. K., Mackinlay, J. D., Shneiderman, B.: Information Visualization. Readings in Information Visualization Using Vision to Think, pp.1-34, Morgan Kaufman Publishers, Inc. 1999.
- [2] 奥村学, 難波英嗣: テキスト自動要約, オーム社, 2005.
- [3] Mackinlay, J., Rao, R., Card, S. K.: The Perspective Wall: Detail and Context Smoothly Integrated. Procs. the ACM Conference on Human Factors in Computing Systems, pp. 173-179, 1991.
- [4] Leuski, A., Lin, C., Hovy, E.: iNeATS: Interactive Multi-Document Summarization. Procs. the 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003), pp.125-128, 2003.
- [5] Kato, T., Matsushita, M., Kando, N.: Expansion of Multimodal Summarization for Trend Information - Report on the First and Second Cycles of the MuST Workshop -. Procs. of the Sixth NTCIR Workshop Meeting on Evaluation of Information Access Technologies, pp.235-242, 2007.
- [6] 加藤恒昭, 松下光範: 時系列情報の抽出と可視化に基づく情報アクセスのためのマルチモーダルインタフェース—情報編纂の基盤技術に向けて—. 人工知能学会論文誌, Vol.22, No.5, pp.553-562, 2007.
- [7] 松下光範, 加藤恒昭: 言語情報と数値情報の相補的利用を目指した可視化手法, 第21回人工知能学会全国大会, 3H8-3, 2007.

### 著者紹介



#### 加藤 恒昭 (かとう つねあき) :

1959年生まれ。1981年東京工業大学工学部電気電子工学科卒業、1983年東京工業大学大学院総合理工学研究科電子システム専攻修士課程修了。同年、日本電信電話公社(現NTT)に入社。2000年より、東京大学大学院総合文化研究科言語情報科学専攻准教授、現在に至る。質問応答、マルチモーダルコミュニケーション、語彙意味論に関する研究に従事。工学博士。電子情報通信学会、情報処理学会、言語処理学会、人工知能学会、ACL各会員。



#### 松下 光範 (まつした みつなり) :

1993年大阪大学工学部精密工学科卒業。1995年大阪大学大学院基礎工学研究科物理系専攻制御工学分野博士前期課程修了。同年4月、日本電信電話株式会社入社、現在に至る。2006年より名古屋大学大学院情報科学研究科メディア科学専攻客員准教授。博士(工学)。自然言語理解、情報可視化、ヒューマンコンピュータインタラクションに関する研究に従事。情報処理学会、人工知能学会、日本知能情報ファジィ学会、日本バーチャルリアリティ学会、ACM各会員。