

## 動向情報の要約と可視化に関するワークショップの提案

加藤 恒昭<sup>†</sup> 松下 光範<sup>††</sup> 平尾 努<sup>††</sup>

<sup>†</sup> 東京大学大学院 総合文化研究科 言語情報科学専攻

〒 153-8902 東京都目黒区駒場 3-8-1

<sup>††</sup> 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所

〒 619-0237 京都府相楽郡精華町光台 2-4

E-mail: †kato@boz.c.u-tokyo.ac.jp, ††{mat,hirao}@cslab.kecl.ntt.co.jp

**あらまし** 動向情報は、製品価格や内閣支持率の変化等、時系列情報に基づき、それを総合的にまとめ上げることで得られるものである。このような動向情報の効率的な提供には、文章だけでなくグラフなどの視覚的手段を利用し、それらを協調させることが必要となる。本稿では、複数文書に分散した様々な動向情報を文章や図表で要約・可視化するという研究課題を提案し、その処理の枠組みを示す。加えて、この課題の研究に有益であろうコーパスについて説明し、これを共通の研究素材とし、動向情報の要約と可視化への関心を共有する研究者によるワークショップを提案する。

**キーワード** 動向情報, 自動要約, 情報可視化, 複数文書要約, マルチメディアプレゼンテーション

## Proposal of a Workshop on Summarization and Visualization of Trend Information

Tsuneaki KATO<sup>†</sup>, Mitsunori MATSUSHITA<sup>††</sup>, and Tsutomu HIRAO<sup>††</sup>

<sup>†</sup> The University of Tokyo Language and Information Sciences

Komaba 3-8-1, Meguro-ku, Tokyo 153-8902 Japan

<sup>††</sup> Nippon Telegraph and Telephone Corp. NTT Communication Science Labs.

Hikaridai 2-4, Seika-cho, Soraku-gun, Kyoto 619-0237 Japan

E-mail: †kato@boz.c.u-tokyo.ac.jp, ††{mat,hirao}@cslab.kecl.ntt.co.jp

**Abstract** Trend information is obtained by synthesis and organization of series of temporal information such as transitions of a product price and a degree of public support for a cabinet. Effective communication of trend information should employ as its media not only text but also visual ones such as charts, and use those in a cooperative manner. In this paper, a research theme is proposed, that allows trend information scattered in multiple articles to be gathered, summarized, and provided in linguistically and/or visually. We show a framework to accomplish this research and explain a corpus useful for that purpose. We also propose a workshop on this research on summarization and visualization of trend information in which the researchers share this corpus as a common material.

**Key words** Trend Information, Text Summarization, Information Visualization, Multiple Document Summarization, MultiMedia Presentation

### 1. はじめに

ある商品の価格や売上の状況、ある会社の業績状況、内閣や政党の支持状況等を動向情報と呼ぶ。動向情報は、幾つかの統計量の時系列データを基として、その変化を通時的にとらえつつ、それらを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ものである。加えて、多くの動向は、製品シェアの動向のように複数の会社に関係したり、地方毎の土地価格

の動向のように空間的な広がりを持つたりするので、それは単なる1次元の時系列情報の要約ではなく、それに関わる複数主体や空間の軸を含んだものとなる<sup>(注1)</sup>。

本稿では、このような動向情報への関心、例えば、「去年から今年にかけてガソリンの値段ってどう動いていますか」「98

(注1) : やや拡大解釈であるが、本稿ではこの主体や空間の軸を中心とする共時的な状況の要約も動向情報に含まれるものとして考えていく。

年頃、パソコンの出荷はどんな感じでしたか」「最近、ソニーは業績よいですか」「去年の台風はひどかったですか」「今期の大リーグのホームラン争いはどんな経過でしたか」に、簡潔で平易な文章や視覚的なグラフで、もしくはそれらを協調させたマルチメディアプレゼンテーションで答えるための、動向情報の要約と可視化という課題について述べる。まず、この課題が、広い意味での情報アクセスに関する技術を様々な点で牽引することを示す。次に、現在作成中の、この課題の研究に有益であろうコーパスの設計方針と仕様を説明する。最後に、このコーパスを共通の研究素材とし、動向情報の要約と可視化への関心を共有する研究者によるワークショップを提案する。

## 2. アプローチ

動向情報の要約と可視化のための枠組みの一例を図1に示す。「98年頃のパソコン業界の動向」が利用者の関心であるとする。以下のように処理が進められる。ある種のオントロジを用いて、パソコン業界の動向が、パソコンの国内出荷台数、国内出荷額、それらのメーカー別シェアという統計量によって説明されることが得られ、更にそれら統計量の関係が把握される。次に、新聞記事のような大規模テキスト集合や白書のような数値データの集まりから、パソコン業界の動向や国内出荷台数等の統計量に関する情報が文書単位で検索・収集される。テキスト集合として新聞記事を考えれば検索されたそれぞれの記事の一部のみが利用者の関心や必要な統計量に関連すると予想されるので、その部分、つまり、パソコン業界の動向についての全般的なコメント「昨年夏以降、絶好調が続き」「右肩上がりの急成長を続けてきたパソコン市場も、景気低迷によって踊り場にさしかかっている」<sup>(注2)</sup>や、それぞれの統計量に関する具体的な値の変化への言及「上半期の国内出荷は前年同期比34%増の439万1000台、半期ベースで過去最高を記録した」等を抽出し、それらの要約が生成される。ここで要約という場合、必ずしも文章による要約に限るものでなく、統計量の変化を描画したグラフによるものも含まれる。また、必要に応じて白書などの数値データも参照される。最後に、これらの要約が、統計量間の関係、例えば、「低価格化が進むと出荷台数の伸びほど出荷金額は伸びない」等を考慮して、総合的にまとめ上げられる。図では、「レポート」という表現を用いているが、文章のみ、グラフのみの要約や、文章を中心にそこからグラフを参照しているもの、グラフを中心にそこにコメントが注釈付けされているもの等、様々な形式が可能である。

このような枠組みにおいて、最初の「関連する統計量とそれらの関係の抽出」は、与えられた関心、例えば、企業業績や業界動向に対して答えるためにはどのような統計量に言及しなければならないか、それらの統計量はお互いにどういう関係にあるか等の知識をどのように蓄積し用いるか、加えてどのように構築・獲得するかという知識工学指向の課題であるので、その一般的な解決はオントロジ工学、セマンティックWeb等の分野の成果の蓄積を待ちたい。次の「情報の収集」については、現

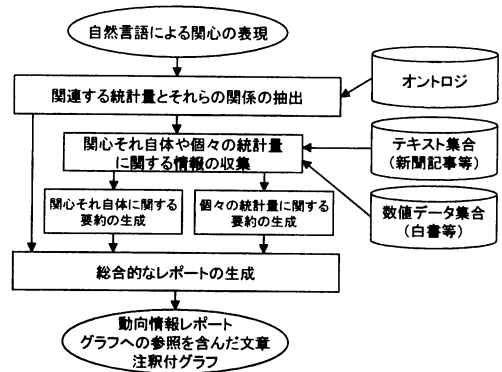


図1 動向情報の要約と可視化の枠組み

状の情報検索技術が充分その役割を果たすものと考えられる。

このように考えると、動向情報の要約と可視化へアプローチするためには、複数文書に分散した特定のトピック、特にある統計量に関する情報を文章あるいは視覚的なグラフとして要約する技術、それら個別のトピックに関する要約を総合的にまとめあげてひとつのレポートを生成する技術が必要となる。

ここで、個別の統計量に関する情報は、白書等の数値データのみから得られるものだけでは不十分であることを指摘しておく。動向情報として必要なのは、「客観的」な値の推移ではなく、それらをどう見るか、それらを用いて何を伝えるかという判断に基づいて抽象化され、強調点を取捨選択された要約である。文章はもちろんグラフもこのような数値の背後にある作成者の意図や関心、達成したい目的によって変化することが知られている[2][4][8][12]。そのための情報は、「純粋」な数値データから得られるものではなく、むしろ、新聞記事のような、判断によって取捨選択された数値情報とそれへのコメントからこそより適切に得られるものである。

そのような関心を伴うとここで、例えば、「原油は上昇、なぜかガソリンは下落」と述べたいのであれば、原油価格とガソリン価格を同じ時間軸のグラフの上に重ねて描き、その傾向が分かるように適切な価格の軸を設定するという可視化が可能になる。また、台風の上陸の動向であれば、どの時期に多かったかに関心があれば時間軸上のヒストグラムを用いるし、どこに多かったかに関心があれば日本地図上に上陸位置をプロットする手法をとる等の選択が可能になる。

## 3. 位置付け

情報検索、質問応答、自動要約等、情報アクセスのための様々な技術が研究されている。これらは、利用者の関心や興味を満たす情報を利用者が最も利用しやすい形で提供するための技術と考えることができる。そのような位置付けであれば、情報可視化も利用者が必要な情報を理解しやすい形式で提供するための技術として同列に議論されることに問題はないだろう。特に、マルチメディアプレゼンテーションの場面では、これらの技術は協調してよりよい情報提供を行うために用いられる[7]。これらの

(注2)：以下、引用はすべて1998年1999年の毎日新聞の記事から行っている。

情報アクセス技術において、動向情報の要約と可視化は、以下のように位置づけることができる。

質問応答技術は自然言語による質問に多量の文書を用いて回答するためのものであるが、その関心は、TREC [16] や QAC [5] を中心とした factoid 質問に関するものと AQUAINT [1] のような多量のコンテキストを背景とした高度な問題解決に関わるものと二分している。factoid 質問を対象とした質問応答技術でもその対話的場面での利用を前提とすれば利用者のより大きな関心に対応できる [9] が、特定の人物や事件に関する情報を収集するような場合に較べて、動向情報は factoid 質問に展開しづらいという印象がある。生年月日や出身地、現在の役職などの質問を繰り返すことである人物のプロファイルが得られるのに対し、特定の時点の統計量の値についての質問を繰り返しても動向は得られず、動向情報については本質的に要約的要素を切り離すことができない。このような課題は factoid 質問を対象とした質問応答技術を一步進めることになると考えられるし、AQUAINT 程には多量のコンテキストを必要としないので、分野に依存しないという質問応答技術の利点を維持したままでの技術の発展につながることを期待できる。

自動要約技術にとって、動向情報を扱うことは大きくふたつの意味を持つ。第一に、動向情報は複数の文書に分散しているので、その要約は複数文書要約 [10] となる。複数文書要約は多くの研究者から関心が向けられており、自動要約の評価型ワークショップである DUC [3]、TSC [15] もそれを主たる対象としているものである。加えて、対象となる複数の文書全体を要約するのではなく、それらの中に点在する、利用者の関心に関わる部分（ここではある統計量に関する記述）のみを要約するという ‘user-focused’ な要約であり、この点も最近の研究トピックとして注目されている [11]。第二に、現在の複数文書要約は、重要な文や節を抽出する処理とそれらから重複を除いて整合的な要約を構成する処理とからなるのが普通であるが、前者については一定の成果が既に得られており、後者の問題がクローズアップされつつある [6]。動向情報は、基本的に時系列情報であり、過去に関する言及が多いことから重複が多い一方で、数値情報ということからそれらの整合性は比較的取りやすいため、後者の重複除去や整合性確保に関する技術のための適当なテストベッドになると考えられる。

情報可視化技術では様々な情報を視覚的に表現することを試みられている [13] ので、時間軸を中心とした統計量の変化を視覚化するという課題は、ある意味で基本的すぎると考えられるかもしれない。しかし、動向情報の可視化は少なくとも3つのインパクトを可視化技術に与える。第一に、時間軸やそれに空間軸等を加えた複数次元のデータを概観することは情報可視化の主たる一分野であるから、動向情報という広い種類のデータに対する技術の確立は可視化技術の共通基盤となる可能性を持つ。加えて、このような観点での可視化は背後にある新聞記事集合と併せるとインフォメーション・アウトライニング [14] の側面もあわせ持った興味深いものである。第二に、従来の可視化技術が多量のデータを捨象要約して表現するという方向であったのに対し、テキスト情報のみから動向を可視化する場

は、少量の不完全な情報を対象として、推論等により情報を補間することが必要になる。この方向の可視化はこれまであまり試みられていない。第三に、動向情報の可視化は、データや事実のみではなく意見や解釈を含んだ情報の可視化である。描くべきなのは、トピック（パソコン業界の状況）についてのグラフではなくより望ましいとされるメッセージ（パソコン業界では昨年夏以降、絶好調が続いてる）についてのグラフである [17]。既に述べたように作成者の意図や関心、伝達したいメッセージがグラフの形式に影響を与えることは多くの研究で報告されており、可視化技術の中心的な課題である。

#### 4. 動向への関心とその情報

具体的に研究を進めるにあたっては、動向への利用者の関心をもう少し厳密に定める必要がある。利用者の関心の「切り取り方」として次の2種類が考えられる。

**統計量（の集まり）や出来事のクラスに着目した切り取り方**  
ガソリン価格に関する動向に関心を持ち、ガソリン全国平均店頭価格とドバイ原油価格というふたつの統計量に着目したレポートを求める、台風上陸という出来事のクラスに関心を持ち、上陸の日付や場所をまとめた今年の台風上陸の傾向に関するレポートを求める等の場合。

**関心のある主体や対象物に着目した切り取り方**  
ソニーの動向、小渕内閣の動向のように、統計量というよりも主体や対象物に着目して、それに関するレポートを求める場合。実際には、一般にソニーの動向はソニーの業績動向であり、関連する統計量の集まり（売上高、営業利益、経常利益等）があるので、この場合は、上の統計量に着目した切り取りを更に関心のある主体で絞り込んだ場合と考えることができる。つまり、すべての企業の経常利益に関心があるわけではなく、ソニーのそれのみ関心があることになる。また、大リーグのホームラン争いへの関心であれば、ホームランという出来事のクラスすべてに関心があるわけではなく、マグワイアとソーサのホームランに関心が限定されることになる。

もちろん、どこまでを主体や対象物と考えるかには恣意性がある。たとえば、ガソリン価格への関心は統計量に関するものとしたが、生活必需品の価格動向ということで、食品や衣服と並べられるのであれば、価格が統計量であり、ガソリンの位置づけは対象物ということになる。これは利用者の捉え方の問題ということもできるが、その捉え方によって生成されるレポートも変化するはずである。

上記のように示された関心に関する情報、つまり、要約や可視化の素材は、大きく分けて以下のいずれかの形をしている。

**それぞれの統計量のある時点（時間幅）での値に関する記述**  
ガソリン価格の動向やソニーの業績動向等の場合は、個々の情報は、「ガソリン価格（レギュラー1リットル）は全国平均で92円となり、前週の平均に比べ1円上昇した」「ソニーは7日の1998年3月期決算発表で、関連会社などを含む連結段階の経常利益が4537億4900万円（前期比45・2%増）に達し」のような、ある時点でのある統計量の値の記述と、加

えてそれと過去の値との比較や評価からなっている。

**ある出来事に関する記述** ある年の台風の動向やホームラン争いの動向の場合、中心となるのは、「中型で並の強さの台風10号は、17日午後4時半ごろ鹿児島県枕崎市に上陸、同日夜には高知県宿毛市付近に再上陸して進み」「25日にはソーサは66号を放ちマグワイアを抜いてトップに立った」のような、個々の出来事の記述であり、その年にどこに幾つの台風が上陸したかという統計的な記述は中心的ではない。ただし、「54号を放ち、マリスの61本の本塁打記録にあと7本と迫った」のように出来事の記述と統計量に関する記述が混在する場合もある。

動向情報の要約と可視化は、これらの情報を抽出し解析し、それらを総合的にまとめ上げることによってなされる。

## 5. 動向情報コーパス

前節の観察に基づき、動向情報の要約と可視化に関する研究のために、次のようなコーパスを作成中である<sup>(注3)</sup>。テキスト集合を毎日新聞1998年、1999年の2年分とし、動向情報として15のトピックとそれに関連する統計量や出来事のクラス(各トピックについて3つ前後)を選んだ。これらについて、以下の情報を作成あるいは収集中である。

- それに関する情報を含んだ記事のリスト(トピック毎)
- それら記事への注釈(注釈仕様は後述する)
- 統計量の変化や出来事に関する文章による要約
- 同時期の統計量の変化や出来事に関するグラフ

文書による要約は、収集された記事全体について、統計量毎に200文字程度の長さのものを作成する。グラフも統計量毎であるが、記事中の情報の描画ではなく、統計DB等の情報を用いて描画したり既存グラフを収集したりすることで揃えていく。このため一部の統計量については欠落する可能性がある。

選択したトピックとその統計量あるいは着目する出来事を図2に示す。トピックの後の括弧内に示したのは、収集された記事数である。

このコーパスとして提供される情報は、図1に示した動向情報の要約と可視化の枠組みにおいて、「情報の収集」によって得られた記事のリスト、および、「個々の統計量に関する要約の生成」の出力として得られる文章およびグラフによる要約に相当するものである。ただし、グラフについては記事中の情報から作成したものでないこと、それを通して伝えたいことが異なっている可能性があることから必ずしも適切な要約であるとは限らない。また文章による要約もひとつの例であることはいうまでもない。

加えて、各記事への注釈は、要約における重要文の抽出結果、およびそれに対する意味処理結果を提供するもので、要約の言語処理に関心を持つ研究者にとっては処理の中間結果として参照できるものであるし、情報の可視化や要約における後半の整合性確保等に関心を持つ研究者にとっては、その入力データと

(注3)：以下、本稿で述べるのは仕様の現状であり、今後、若干変更される可能性がある。

ガソリン価格 (20)
レギュラーガソリン全国平均店頭価格、トバイ原油価格
ソニー(の連結業績) (9)
連結売上高、連結営業利益、連結経常利益、連結最終利益
パソコン業界(出荷状況) (20)
国内出荷台数、国内出荷額、メーカー毎の出荷台数シェア
ビール(および発泡酒)業界 (21)
出荷数量、メーカー別出荷数量、メーカー毎の出荷量シェア、
ビール発泡酒総市場における発泡酒の割合
通信機器の利用 (11)
(固定電話ポケベル、PHS、携帯電話)の加入台数
人口 (12)
総人口、男女別人口、年代別人口(子供、老年・高齢者)、
出生率
失業者 (14)
完全失業者数、事由別完全失業者数、完全失業率、
男女別完全失業率
台風 (15)
台風の上陸場所・日時(出来事)
大リーグのホームラン争い (13)
(マグワイヤ、ソーサ)のホームラン(出来事)、
(マグワイヤ、ソーサ)の年間ホームラン数

図2 トピック、統計量、出来事の例

して利用できるものとなっている。注釈の主な仕様を以下に示す。また、注釈と要約の例を付録に示す。

**unit 要素** 指定された統計量や出来事に言及している部分。原則として文を単位とする。言及している統計量(stat)や出来事(event)は属性として記述される。また、その情報が事実でなく予測である場合はtype="pros"が付与される。すべての注釈はこの要素について行われる。

**del 要素** unit 要素の一部であるが、情報の出典や根拠あるいは関連する出来事など、統計量の値や変化に関する直接の情報でない部分や対象となる出来事の記述ではない部分。原則として節を単位とし、統計量を説明する連体修飾等はこの要素とならない。

**ins 要素** unit 要素の情報をそれだけで完結したものとするために必要な補完された部分、主に統計量名やパラメータ名に適切な助詞を加えたもの。補完された後の表現も日本語とし適格である。

**name 要素** 統計量の名前。どの統計量であるかはstat属性で示されるが、これを含むunit要素のstat属性の値が一意である場合は省略される。また、「ガソリンの価格は全国平均で」のように統計量の名前に相当する部分が分かれてあるいは一部のみが出現する場合があります。その際は意味的に主たる部分にpart="head"、それ以外にpart="foot"の属性を付与している。この例では、「<name part="head">ガソリンの価格</name>は<name part="foot">全国平均</name>で」となる。

**val 要素** 統計量の値。stat属性についてはname要素と同じ。

rel 要素 統計量の値そのものではないが、その値の差や比、順位等の相対値を表現する部分。「高値」等の具体的でない表現は含まない。stat 属性については name 要素に同じ。

date 要素 「10日」「今月」「昨年」等の時刻の表現。gra 属性として、日、月、年等その時刻表現の粒度が、abs 属性として、その表現が指示する日付（年の粒度の場合は西暦4桁、日の場合は8桁）が示される。例えば、1998年6月17日付の記事での「先週」は、「<date gra="週" abs="19980610">先週</date>」とされる。日付の表現であっても統計量に関する情報の一部でない場合はこの要素とならない。

dur 要素 「4カ月（ぶり）」等の期間の表現。date 要素と同様に gra 属性を持つ。「第一四半期」のように統計量に対して時点として振る舞うものは期間であっても date 要素となる。

par 要素 日付以外の統計量のパラメータ、出来事をおこした主体やその場所等、出来事の一部となる事物。シェアにおける会社名や台風上陸の場所や台風の号数等がこれにあたる。ただし、利用者の関心として指定されたもの、ソニー、ソーサ、マグワイヤ等は特別に扱われ、topic 要素とされる。

この他に、統計量に関する付帯条件である cond 要素、「同」等の参照表現についての pro 要素等がある。

## 6. ワークショップの提案

動向情報の要約と可視化に関する技術について、共通の素材を用いて、協調的かつ競争的に研究を進めていくワークショップを提案したい。前節で述べたコーパスは遅くとも2004年末には完成するので、これを共通の場での研究成果報告を条件に広く利用していただくことを考えている。研究成果（もしくは経過）の報告時期は2006年1月から3月頃を計画しており、形式としては、学会のシンポジウムや大会に併設されたワークショップの利用、あるいはそれに合わせてクローズドな報告会の開催を考えている。共通のデータを用いて、緩い意味で共通の課題に取り組むことによる議論と研究の活性化、ツールやコーパス類の蓄積をワークショップの目的とし、いわゆる評価型ワークショップで行われている客観評価の実施は、少なくとも今回は考えない。コーパスに加えて、数値情報可視化のためのグラフ描画のツール等についてもワークショップ内での共有を考えていきたい。なお、現状では、コーパスを利用するためには毎日新聞 CD-ROM 1998年、1999年版の購入が必要であるが、本コーパスに含まれる記事について、ワークショップ参加者を対象に研究目的での無償使用を毎日新聞社と交渉中である。

## 7. おわりに

複数文書に分散した様々な動向情報を文章や図表を用いて要約・可視化するという研究課題を提案し、その概要を示した。この課題は、様々な情報アクセス技術、質問応答技術、自動要約技術、情報可視化技術を牽引するものである。この課題に関する研究で利用可能なコーパスを紹介し、それを共有し、共通する課題に取り組むことで議論と研究の活性化を図るワーク

ショップを提案した。

## 謝 辞

本研究は、NTTと東京大学との産学連携共同研究、ならびに国立情報学研究所のNTTと東京大学との公募型共同研究によって支援されています。ご支援をここに感謝いたします。国立情報学研究所の神門典子氏、東京工業大学の奥村学氏、広島市立大学の難波英嗣氏からは、貴重なコメントをいただきました。あわせて、ここに感謝いたします。

## 文 献

- [1] AQUAINT Home Page: Advanced Question & Answering for Intelligence  
<http://www.ic-arda.org/InfoExploit/aquaint/>. 2003.
- [2] Sandra Carberry, Stephane Elzer, et al. "Extending Document Summarization to Information Graphics" *Proceedings of the ACL-04 Workshop: Text Summarization Branches Out*, pp. 3 - 9, 2004.
- [3] Document Understanding Conferences Home Page  
<http://duc.nist.gov>, 2004.
- [4] Massimo Fasciano and Guy Lapalme. "Postgraphe: A system for the generation of statistical graphics and text" *Proceedings of 8th International Workshop on Natural Language Generation*, pp. 51 - 60, 1996.
- [5] Jun'ichi Fukumoto, Tsuneaki Kato and Fumito Masui. "Question Answering Challenge (QAC-1) An Evaluation of question answering tasks at the NTCIR workshop 3" *AAAI 2003 Spring Symposium. New Directions in Question Answering*, pp. 122 - 133, 2003.
- [6] Tsutomu Hirao, Takahiro Fukusima, et al. "Corpus and Evaluation Measures for Multiple Document Summarization with Multiple Sources" *Proceedings of 20th International Conference on Computational Linguistics*, pp. 535 - 541, 2004.
- [7] 加藤 恒昭. "マルチメディアプレゼンテーションの自動生成に向けて - 自然言語生成からマルチメディア生成へ -" *情報処理*, Vol. 38, No. 12, pp. 1049 - 1056, 1997.
- [8] 加藤 恒昭, 松下 光範. "グラフを用いた探索的データ分析のためのマルチモーダル対話処理" *電子情報通信学会論文誌, D-II* Vol. 187 No. 5, pp. 1142 - 1152, 2004.
- [9] 加藤恒昭, 福本淳一, 榊井文人, 神門典子. "質問応答技術は情報アクセス対話を実現できるか" *情報処理学会研究報告*, 2004-NL-162, pp. 145 - 150, 2004.
- [10] 難波 英嗣, 奥村 学. "ここまで来たテキスト自動要約" *情報処理*, Vol. 43, No. 12, pp. 1287 - 1294, 2002.
- [11] 奥村 学. "テキスト自動要約" *情報処理*, Vol. 45, No. 6, pp. 574 - 579, 2004.
- [12] Steven F. Roth and Joe Mattis. "Data characterization for intelligent graphic presentation", *Proceedings of Conference on Human Factors in Computing Systems*, pp. 193 - 200, 1990.
- [13] Robert Spence. *Information Visualization*, Addison-Wesley, 2001.
- [14] 武田 浩一, 野美山 浩. "テキスト情報の可視化を利用した情報検索" *情報処理*, Vol. 41, No. 4, pp. 343 - 350, 2000.
- [15] Text Summarization Challenge Home Page  
<http://lr-www.pi.titech.ac.jp/tsc/>, 2003.
- [16] Ellen M. Voorhees and Dawn M. Tice. Building a Question Answering Test Collection *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 200 - 207, 2000.
- [17] Gene Zelazny. *Say It with Charts*, McGraw-Hill, third edition, 1996.

## 付 録

### 注 釈 例

#### 990617020 ガソリン価格

<unit stat="レギュラーガソリン全国平均店頭価格"><del>石油情報センターが16日発表した給油所石油製品市況調査によると、</del><date gra="週" abs="19990617">今週</date>調査の<name part="head">ガソリン価格(レギュラー1リットル)</name>は<name part="foot">全国平均</name>で<val>92円</val>となり、<date gra="週" abs="19990610">前週</date>の<name part="foot">平均</name>に比べ<rel>1円</rel>上昇した</unit>。<unit stat="レギュラーガソリン全国平均店頭価格"><name part="foot">全国平均</name>の上昇は<date gra="旬" abs="19990201">2月上旬</date>以来、<dur gra="月">4カ月</dur>ぶりだ</unit>。

#### 980508068 ソニー

<unit stat="連結経常利益"><del>消費低迷や半導体市況の悪化などで電機メーカー各社の業績悪化が相次ぐ中、</del><topic>ソニー</topic>は7日の<date gra="四半期" abs="199803">1998年3月期</date>決算発表で、関連会社などを含む<name>連結段階の経常利益</name>が<val>4537億4900万円</val>( <date gra="四半期" abs="199712">前期</date>比<rel>45・2%</rel>増)に達し、<rel>過去最高益</rel>を更新したことを明らかにした</unit>。昨年発売した平面ブラウン管テレビ「ベガ」や、ノート型パソコン「バイオ」など多様なヒット商品に支えられ“ひとり勝ち”した格好だ。

<unit stat="連結売上高, 連結営業利益"><name stat="連結売上高">連結段階の売上高</name>は<val stat="連結売上高">6兆7554億円</val>( <rel stat="連結売上高">19・3%</rel>増)、本業のもうけを示す<name stat="連結営業利益">営業利益</name>も<val stat="連結営業利益">5202億1000万円</val>( <rel stat="連結営業利益">40・5%</rel>増)と、<name stat="連結売上高, 連結営業利益, 連結経常利益">史上最高</name>だった</unit>。

#### 980122071 パソコン業界

<unit stat="メーカー毎出荷台数">メーカー別では、<dur gra="月" abs="1997">年後半</dur>から<par>ソニー</par>のノートパソコンが急伸</unit>。<unit stat="メーカー毎出荷台数シェア"><par>NECなど昨年の上位5社</par>の<name>シェア</name>は同<rel>3・1ポイント</rel>低い<val>82・7%</val>となった</unit>。<unit stat="メーカー毎出荷台数シェア"><par>アップルコンピュータ</par><ins>の<name>シェア</name></ins>は経営危機のニュースなどで<dur gra="月" abs="1997">年前半</dur>に苦戦し、<pro ref="<date gra="年" abs="1996">前年</date>比">同</pro><rel>2・8ポイント</rel>下落して<val>7・8%</val>と<val>10%の大台</val>を割り込んだ</unit>。

#### 981018004 台風の上陸

<unit event="台風上陸">中型で並の強さの<par>台風10号</par>は、<date gra="時" abs="19980917">17日午後4時半ごろ</date><par>鹿児島県枕崎市</par>に上陸、<date gra="時" abs="19980917">同日夜</date>には<par>高知県宿毛市付近</par>に再上陸して進み<del>、西日本を縦断した</del></unit>。

#### 980928250 大リーグのホームラン争い

<unit stat="年間ホームラン数">シカゴ・カブスの<topic>サミー・ソーサ</topic>外野手(29)</topic>は最終戦では5打席2安打で快打がなく、<name>本塁打数</name>は<topic>マグワイア</topic>に<rel>4本</rel>差の<val>66本</val>にとどまっている</unit>。

ソーサは、マグワイアと最後までホームラン競争を演じ、野球の興奮をかきたてた。<unit event="ホームラン" stat="年間ホームラン数"><date gra="日" abs="19980925">25日</date>には<ins><topic>ソーサ</topic></ins><par>66号</par>を打ち<topic>マグワイア</topic>を抜いて<rel>トップ</rel>に立ったが、<date gra="時" abs="19980925">45分後</date>に<topic>マグワイア</topic>が打ち、並ばれた</unit>。

先頭の9桁の数値は毎日新聞の記事番号である。このうち、先頭の6桁が記事が掲載された年月日を示している。

### 要 約 例

#### レギュラーガソリンの全国平均店頭価格

1リットル当たりのレギュラーガソリンの全国平均店頭価格は92年12月の125円をピークに下がり続け、97年後半には100円台を割っていた。98年4月には、さらに下がり94円となり、その後6月から11月は92円の超安値が続いた。99年3月には91円、5月には90円の最安値を更新した。しかし6月には2年半ぶりの前月比プラスに転じ92円、その後も上昇を続け8月には94円、9月には最安値から5円の上昇を見た。

#### パソコンのメーカー毎の出荷シェア

97年度のメーカー毎のパソコン出荷量シェアは、1位のNECが30・9%、2位の富士通が24・1%で順位は前年と変わらず。アップルコンピュータは7・8%と10%の大台を割り込んだ。NECなど上位5社の合計シェアは3・1ポイント下落の82・7%となった。98年は1位NECが28・6%に下がり、富士通、日本IBMを含め上位5社はいずれもシェアを下げた。99年の後半期にはNEC、富士通、日本IBMの上位3社がシェアダウンした一方、東芝がアップルコンピュータを抜いて4位浮上した。

#### マグワイアのホームラン

ホームランのシーズン最多記録に挑むカージナルスのマーク・マグワイア一塁手(34)は98年8月26日マーリンズ戦で54号となるホームランを打った。その後114試合目にシーズン最多記録に並ぶ61号、9月8日にはシーズン最多記録となる62号を打った。さらに9月19日には64号、20日に65と記録を更新、9月25日には一時ソーサ選手にトップを譲ったが、その40分後には66号を打ち、トップに返り咲いた。最終的にマグワイア選手のホームランは70本となった。