特集論文

情報編纂研究促進のための試み

Attempts for fostering researches on information compilation

松下 光範
Mitsunori Matsushita
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*1
NTT Communication Science Labs., NTT Corp.
mat@cslab.kecl.ntt.co.jp, http://www.kecl.ntt.co.jp/csl/msrg/members/mat/index-j.html

加藤 恒昭
Tsuneaki Kato
東京大学大学院 総合文化研究科
Graduate School of Arts and Sciences, University of Tokyo
kato@boz.c.u-tokyo.ac.jp

keywords: information compilation, multimodal summarization, trend information

Summary

Information compilation is a novel research topic which aims to compile various information intelligently, and to make it easy to comprehend/access. The information compilation is emphasized by two characteristics. The first one is a cross-modalitiness achieved by cooperation between non-linguistic information and linguistic information, and the second one is a continuousness in supporting all aspects of information utilization. To foster researches related to the information compilation, we conducted three attempts: installation of a reference model on information compilation, distribution of annotated corpus and a visualization platform as boundary objects, and deployment of an evaluation workshop. The paper describes current efforts of the attempts.

1. ま え が き

近年，様々な情報が電子化されネットワーク上に蓄積されるようになってきている．それに伴い，これらの情報を利用して意思決定や問題解決に役立てる試みが進められている [藤本 05]．しかし，蓄積された情報の量や種類はすでに膨大なものとなっているうえ，時間の経過に伴って更に増加を続けている．そのため，“情報の在り処を見つける” ことを主眼とした検索技術ではユーザの要求に十分に応えることができず，ユーザの意図や関心に応じて適応的に情報を纏め上げ，それへの簡便なアクセスを支援する技術，言うなれば “情報の理解を助ける” 技術が渇望されている．このような背景の下，我々は新聞記事テキストや統計データといった異なるモードの情報を相補的に用いて編纂し，ユーザの情報アクセス行為を容易にする技術（情報編纂技術）の実現を目指している [加藤 06]．

ネットワーク上にはテキストだけでなく音声や画像，動画など様々な種類の情報が存在している．情報編纂技術が目指すのは，これらの情報をユーザの興味や関心に基づいて適応的に再構成し一纏まりの情報として提示すると共に，それをトリガとしたインタラクティブな情報アクセスを可能にすることである．

このような課題に対して，著者らはふたつの取り組みを行っている．ひとつは情報編纂の中心的な技術についての具体的な検討とそれを具現化したシステムの構築である [松下 06, 松下 07, 加藤 07b]．そして，もうひとつは情報編纂という技術分野（研究分野）を確立し，多くの研究者の関心を喚起するとともに，彼らが協調的・競争的に研究を行うことを可能とし，かつそれを加速するための研究者コミュニティおよび研究環境の構築・支援である [加藤 04, Kato 07c]．この情報編纂技術の構築には，これまで比較的疎遠であった様々な技術を創造的に融合することが必要になる．また，後述するようにその裾野も広い．これに取り組むに際して，前者の個人的な取り組みだけでは限界があり，後者の研究者コミュニティによる協調的・競争的な研究は欠かせない．それを促進することは，直接の研究開発に優るとも劣らない重要性があると認識している．

情報編纂技術を促進させるための取り組みの一環として，筆者らは人工知能学会全国大会の近未来チャレンジのテーマのひとつである「情報編纂の基盤技術」を主催している．「情報編纂の基盤技術」は，筆者らが 2006 年度の人工知能学会全国大会で新規チャレンジとして提案して採択され，2007 年度および 2008 年度の 2 度の近未来チャレンジセッションを開催し現在に至っている．

2006 年の新規チャレンジの際に，筆者らが描いた目標達成のための技術的ステップは以下の通りである [加藤 06]．

Step 1. 数値データと言語情報を動向としてまとめ，その動向を言語的視覚的に表現し，それを通じてその

*1 現職: 関西大学総合情報学部

背後にある詳細な情報への柔軟なアクセスを可能にする．ここで扱うのは時間的空間的な広がりを持つ情報と出来事に関する情報が中心である．

Step 2. 次のステップとして，動向を理解・活用する際に必要となる関連情報，すなわち評価・評判，態度・意見など，を対象としてそれらの情報を抽出・可視化するための基盤技術を確立する．

Step 3. さらに画像理解や画像要約を行うために，音声や動画など，非言語情報を取り扱うための技術へと展開する．

本チャレンジではこのようなステップを参加者間の共通了解とすることで，それに向けた個々の要素技術の精緻化・高度化を図るとともにそれら要素技術間の連携を促すことを狙っている．また，その取り組みを支援するための施策を講じることで，研究者・研究機関同士が協調的・競争的に研究を推進するのに適したコミュニティが形成されることを期待している．

本稿では，情報編纂技術を実現するために必要なこれら一連の研究を複数の研究機関で協調的･競争的に行うための試みについてその現状を報告するとともに，それを促進させるための施策について述べる．まず 2 章で，情報編纂技術の特徴について整理し，研究者コミュニティによる協調的・競争的な研究が重要となることを述べ，そのために必要となる 3 つの方法論（施策）を位置づける．第一は参照モデルの設定，第二は境界オブジェクトとしてのタグ付きコーパスと可視化プラットホームの提供，第三は評価なきワークショップから評価ワークショップへの展開である．その後の 3 つの章 (3〜5 章) で情報編纂技術におけるそれらの施策を説明する．そして，6 章でそれらの現状とその評価を述べ，最後に全体をまとめる．

2. 情報編纂技術と取り組み

大量なデータの中から興味深い知見や特徴を見出す技術としてはデータマイニングや知識発見の研究が進められている [Fayyad 96]．これらの技術の主眼は，大規模なデータ集合から興味深い現象や有意義な規則を抽出することである．すなわち，これらの手法はデータ主導の方法論であり，ユーザは得られた結果の価値判断を行うだけで，どのような結果を得るかについての積極的な働きかけはしない．これに対して我々が目指す情報編纂技術では，ユーザが計算機とのインタラクションを通じて積極的に探索過程に関与しそれによって新たな知見や事実を得ると同時に，その得られた知見や事実を解釈することによって生じた新たな興味や関心を探索行為に反映させるという相互作用性 (interactivity) に着眼している．したがって，ユーザの要求に応じて様々な観点から情報を “概観” できるように情報を収集・加工する技術基盤と，それらを起点とした円滑な情報アクセスを可能にする技術基盤とが情報編纂技術の要諦となる．

様々な知的処理を導入することで利用者の情報要求を的確に判断し，収集される情報そのものの質を向上させるという情報検索技術の高度化もひとつの方向であるが，そうではなく，むしろ，情報検索においては既存の技術レベルを前提とし，その結果を編纂することで，利用者の次のアクションを支援する仕組みを考えていく．

この情報編纂技術に関する研究を進めていくにあたり，我々は二つの力点を置いている [加藤 07a].

第一は，様々なモード/メディア/ジャンルを横断して，現実世界の雑多な情報を入力として扱い，その出力においても視覚情報等の非言語情報と言語情報を協調させたマルチメディアプレゼンテーションを活用していくというモード横断性である．この横断性を通じて，これまで分野内で閉じていた技術群の建設的な融合 (例えば言語処理技術と情報可視化技術の融合など) を図り，それを足掛かりとした新しい技術分野の創成を目指している．入力として扱われる非言語情報については大きくふたつに分けて考えている．ひとつは，表データやチャートのようなシンボリックな意味が明確なものであり，もうひとつはビットの集まりとして画像や音声等のようなパタンとして扱うべき情報で，そこからシンボリックな意味を抽出する事自体に，画像理解，音声理解等の技術を必要とするものである．情報編纂では，後者の処理に必要なパタン理解技術は直接の関心の外とし，それを扱う場合も，パタン理解技術によって得られた意味が注釈された形式を前提とする．加えて，研究の進め方としても，シンボリックな意味が明確な表データやチャートを最初の目標としていく．

第二は，情報の理解・概観の支援と情報へのアクセスの支援とを縫い目なく統合することによる情報活用支援の連続性である．これまで，要約と情報アクセスインタフェースとして区別されてきたもの，要約の中でも，全体を眺めるための報知的要約と，情報へのアクセスを行う手がかりとなる指示的要約 [奥村 05] と区別されてきたものをシームレスに扱えるようにすることで，ユーザの情報アクセス行為を円滑化することを狙っている．

後述するように，情報編纂技術という枠組みはそれ以前より著者らがオーガナイザに加わって行っている「MuST: 動向情報の要約と可視化に関するワークショップ (以後，MuST と記す)」の発展である．情報編纂と比較すると MuST は，(1) 処理の対象とする非言語情報として，シンボリックな意味が明確な非言語情報の典型である時系列数値情報等の統計情報を中心に扱っていた，(2) MuST では情報アクセスの起点となる概要生成に重心がおかれ，その後の対話についての議論はやや希薄であった，という違いがある．

我々が MuST において動向情報と呼んでいるものは以下のような特徴を持ったものである．

- 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，情報の間に重複も多く，組織化が重要となる．
- 時系列データや地理的な情報を含むものが多いことに加えて，それ以外にも様々な整理の観点が存在する．
- 統計量への言及等の事実情報に加えて，その解釈，原因の推測，波及効果の予測等が含まれる．
- 新聞記事や blog のようなテキスト情報の他に統計量に関する数値情報を利用する等，情報の形態にとらわれない情報収集が必要となる．

MuST の取組みは情報編纂技術の観点からすれば部分的なものではあるが，動向情報の持つこれらの特徴から情報編纂技術展開の足掛かりとして一定の役割を果たすと思われる．そこで，いわゆる動向情報を情報編纂技術の現在の核として議論を進めている．

一方，近未来チャレンジ「情報編纂の基盤技術」では，1 章で述べた目標達成のための技術ステップを共通了解とし，MuST で扱うような数値情報と言語情報の相補的利用の研究だけでなく，非言語情報まで含めたより広範な技術への展開や新しい適用領域の開拓をも対象にしている．MuST の取り組みが一つの技術課題を深く一貫した形で掘り下げるのに対して，本チャレンジはそれを核とした技術的な広がりを追求するものとなっており，両者を推し進めることで，効率的効果的な情報編纂技術の発展が可能になると考えている．実際，現時点までででも多様かつ興味深い研究が行われている．MuST および近未来チャレンジのこれまでの成果の概略に関しては 6 章で紹介する．

これまでの MuST および近未来チャレンジの取り組みを通じて，コミュニティを形成し研究を促進させるために今後取り組むべき事項も明確になってきている．

上述したように，情報編纂技術の展開には言語処理技術と情報可視化技術のような，今まで比較的疎遠であった分野の技術の融合や，ある研究分野内においても異なる技術と考えられていたものを統一的な視野で眺めることが必要となる．加えて，情報検索技術との連携やパタン理解技術の活用も検討すべきであろう．更には，想定する利用者を個人からグループにまで広げると，コミュニティウェア，CSCW ツールとしての展開も視野に入れる必要が生じてくる．このような展開を志向した場合，個人の研究者，単独の研究組織が独自に問題に取り組むだけでは限界があり，異なる専門性を持つ専門家や研究グループが研究ゴールに関する共通了解の下で関与するコミュニティが形成されることが不可欠になる．したがって，このようなコミュニティの形成を支援する（形成を促す）ことも研究そのものに劣らず重要であると認識している．

このようなコミュニティは “Community of Interests (CoI)”[Arias 00] と位置付けられるものであり，その形成には研究枠組の共有のための参照モデルの設定が必要になる．この参照モデルは，コミュニティの内側の研究者がこれを通じて自分の研究を位置づけることを可能にすると同時に，コミュニティの外にいる研究者に対して，自らの関心がこの枠組の一部となっていることに気づかせる等，コミュニティに研究者を引き寄せる役割も果たすことになる．また，円滑な課題遂行には異なる研究主体間の情報共有/相互理解を助ける境界オブジェクト (boundary object) の設定が必要になる [Fischer 01]．適切な境界オブジェクトを提供することで研究者間の理解も深まり，研究そのものの効率も高まる．更にコミュニティの形態も重要なものとなる．研究ゴールを共有する研究者のコミュニティとして評価ワークショップや shared task の実施が知られるが，それに留まらない柔軟な形態が必要である．以下の 3 章では，情報編纂技術研究のコミュニティのためのこの 3 つの施策について述べる．

3. 情報編纂の参照モデル

参照モデルは研究の枠組みを定め，個々の研究課題を明確にする役割を果たす．以下では，我々が現時点においてコミュニティのメンバに提示している情報編纂の参照モデルについて説明する．

3·1 参照モデルの概要

情報編纂技術では，例えば次のようなユーザの情報アクセス行為を想定している．
- ガソリン価格の動向に関心がある人が，過去数年の傾向を見て今後どうなるだろうかということに更なる関心を抱いたり (将来予測)，特定の時期に大きな変化があったことに気づいて，なぜそうなったかに興味を抱く (原因究明).
- デジタルカメラの購入を検討しているユーザが，blog などの評判情報をまず概観することで画質に関するネガティブなコメントが多いことに気がつき，具体的に何が悪いと言われているかを調べていく (詳細把握).
- 地震予知に興味を持つ人が，過去に発生した大型地震の時空間的な発生状況から，ある時期やある場所 (もしくはそれらの重ね合わせ) に注目する (仮説生成).

これらの例ではいずれも，事前に情報アクセスの方針を持たず，知りたい情報を概観することで具体的な情報アクセスの方針 (e.g., 将来予測, 原因究明) を得ることを想定している．このようなユーザの振舞いは情報検索の分野で提唱されている berry picking モデル [Bates 89] の延長線上にある．ただし，情報編纂技術が目指すのは，情報との単なる encounter[Marshall 05] の支援ではなく，その時点でのユーザの興味に応じて編纂された情報の提供である．例えば，ユーザが興味を持った情報のタイプ (e.g., 予測，原因) や時期に基づく絞り込み，時間的空間的な粒度に応じた情報の集約，重複情報の計量や要約などが編纂された情報となる．したがって，情報編纂の参

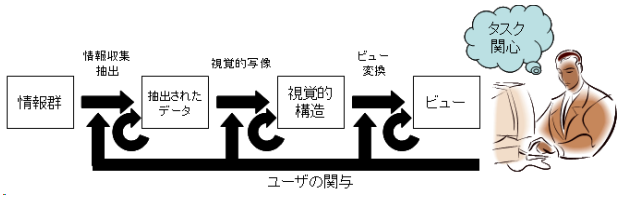


図 1 情報編纂の参照モデル

照モデルは one-shot の編纂情報生成ではなく，ユーザの関与を意識したモデルでなくてはならない．

この点を考慮した情報編纂の参照モデルを図 1 に示す．これは情報可視化の参照モデル [Card 99] を拡張したものである．情報編纂では，まず対象とする情報群からユーザのタスクや関心に基づいて必要十分なデータを絞り込んで抽出する処理が行われる．ここでの処理は，各々の情報の主題抽出・分類，報知的要約文の生成，テキストの時間的空間的なアライメントの他，評判情報であれば何 (e.g., 価格，画質，重量) に関するどんな意見なのか (ポジティブ/ネガティブ) の判定，数値情報であれば粒度変更や集約処理など，ユーザに提供する情報内容の生成に関わるものが含まれる．次に，この処理結果をユーザに提示するために視覚的写像 (visual mapping) を施す．ここでいう視覚的写像は必ずしもグラフなどの可視化表現への写像だけではなく，指示的要約の生成などテキスト表現への写像をも含んでいる．視覚的構造化が施されたデータを，ユーザの興味に合致するように見せ方を決めるのがビュー変換である．ユーザはこの 3 つの過程に関与することで，提示内容をインタラクティブに変更していく．

前述したように，本研究では情報編纂技術の足掛かりとして数値情報と言語情報との連携に着目し，ユーザがそれらの情報にアクセスしたり，その概要を把握したりする際の支援となる技術の実現を目指している．以下ではそれを念頭において図 1 の参照モデルの各段階における処理について述べる．

3·2 情報収集・抽出

情報編纂技術は，様々なメディアの情報をある関心や興味の下で纏め挙げるという点で，マルチモーダル要約と見なすことができる．

テキスト情報の要約には，単一の情報リソースを要約する単一文書要約と，複数の情報リソースを纏めあげる複数文書要約がある [奥村 05]．情報編纂が目指すのは様々なモード/メディア/ジャンルの情報リソース群からユーザの関心に沿った情報内容を抽出し提示することであるため，テキスト情報における複数文書要約に相当する．

一般的な複数文書要約の流れは ① 関連するテキストの自動収集，② 重要文の抽出，③ 冗長性判定，④ 重要箇所の抽出，⑤ 書き換え，⑥ 要約表現の生成である．情報編纂の参照モデルにおける情報収集・抽出処理は複数文書要約における ①〜④ の処理に相当する．

(1) 関連する情報群の自動収集

ユーザから与えられる関心やタスクに基づいて，テキスト情報や数値情報 (将来的には画像や音声，動画などのマルチメデシア情報をも含む) を収集する．この処理に関しては，現在行われている情報検索技術が利用できると考えている．

(2) 主題の抽出と分類，具体化

収集された情報を，意味内容に基づいて断片化し，それらの情報断片の主題を抽出すると共に，編纂の際に利用できるように分類される．例えば原油価格の上昇を伝える記事の中では，しばしば同一記事中に WTI 原油取引価格 (1 バレルあたりの取引価格) や市場でのガソリン販売価格 (1 リットルあたりの小売価格) に関する記述が含まれている．これらを区別しておくことで，ユーザの関心，例えば原油価格の高騰にあるのかそれともガソリン小売価格の上昇にあるのか，に応じて適切な情報を伝えることができるようになる．また，評判情報の場合であれば「何」に関する「どのような」評価かを，数値情報であればそれがどの粒度・単位の値であるかを判別する必要がある．照応解消や相対表現の具体化もこの処理に含まれる．

3·3 視覚的写像

視覚的写像は，複数文書要約における ⑤ と ⑥ の処理に相当する．

情報編纂ではテキストと可視化を組み合わせた提示を行うため，情報リソースとは異なるメディアに変換して提示するほうが効果的な場合がある．例えば，数値情報から全体的/局所的な傾向を判断し，言語表現として提示したり，文書中に含まれる定性表現に基づいて視覚表現を生成したりすることがそれにあたる．このような処理を行うために，メディア変換規則やテンプレートを事前に用意しておく必要がある．

また，3·4 節で後述するように，言語表現と視覚表現の連携を行うためには，異なるモードの情報間のアライメントを行っておく必要がある．例えば，時系列変動のグラフで大きな変動があった箇所を視覚表現で強調しつつ言語情報でその理由や波及効果に言及する場合，視覚表現 (グラフ) で表現される変動箇所と言語表現で表現される情報との間のアライメントが無ければ，言語表現で表現された情報が何に関する言及であるのか分からなくなってしまう．

複数文書要約の冗長性判定と異なるのは，一方のメディアで言及される情報が他のメディアでも強調のために言及する可能性がある点である．例えば，上記の例で，理由や波及効果のような関連情報ではなく，言語表現でもその箇所が大きな変動箇所であると言及することで，ユー

ザに注目を促すような提示効果が得られる*2.

3·4 ビュー変換，ユーザの関与

数値情報と言語情報は，各々単体でも情報を伝達する有効な手段であるが，それらを組み合わせて詳細な情報 (元の記事や白書など) へアクセスする際のインタフェースとして機能させることでより効果的な情報提示が可能になる．そのためには，利用者の意図・関心の違いやその変化に応じて，様々な詳細度でインタラクティブに情報を眺められることが必要である．特に，予め明確な意図や関心があって情報にアクセスするのではなく，大量のデータの中から役に立つ情報を探索的に見つけ出すような分析の場面では，計算機とユーザとのインタラクションが重要な役割を担う．図 1 のビュー変換とユーザの関与がこれに関わる．

グラフなどの視覚表現の場合，表示領域の移動やズームなどの操作，異なる粒度への変更などをユーザに許すことでその要求に応える事が可能となる．これに対して言語情報の場合は，何を内容に含めるかについて観点の異なる要約を提供したり，その要約が生成された際の元情報へのアクセスを許したりすることによりその要求に応えることが可能となる．

また，時系列数値情報に基づく視覚表現と言語情報を併用した情報アクセス行為では，ユーザは視覚表現 (グラフ) を眺めることにより特徴的な箇所に気づき，それに関する知見を得ようとして言語情報を参照する場合と，言語情報から気になる箇所を見つけ，それがグラフのどの部分に対応しているかを参照する場合が想定される．そのため，これらの行為を支援する情報提示では (A) 視覚表現を介して関連する言語情報にアクセスするインタラクションと，(B) 言語情報から視覚表現の対応箇所にアクセスするインタラクションの双方向のインタラクションを提供することが望ましい．

2 章 で述べたように情報編纂では概要把握から詳細情報獲得までの連続性が重要であり，その過程を通じて “ ユーザの観点から情報群を再構成 (編纂) する” ことを支援する技術を目指している．Shneiderman の Visualization Seeking Mantra [Shneiderman 98] によれば，視覚的に情報を探索するユーザの行為は “Overview first, zoom and filter, then details on demand” である．これは，ユーザの探索行為が一般に「全体」から「部分」へ向かうことを示唆しているが，複数モード (視覚表現と言語情報) を併用する情報アクセスの場合，この「全体」と「部分」は状況によって異なってくる．例えば，ユーザがテキスト一覧中のある記事に着目し，一連の流れの中におけるその記事の位置付けを知るためにグラフ上の関連する値やその周縁部を参照する場合は，言語情報が「全体」の役割を，またそれに関連するグラフのある領域が「部分」に相当する．逆にグラフが「全体」，言語情報が「部分」という役割を果たす場合もある．したがって，このようなモードを超えた「全体」と「部分」の相互参照を可能にしてユーザがそれらの関係を容易に把握できるようにすると同時に，それらの切り替えを容易にするインタラクションの枠組をシステムが提供しなくてはならない．

*2 異なるメディアの間でどのような連携を行うかについてはタスクの性質やユーザのその時点での関心に依存するため，連携規則の整備と適用も必要になる．

4. 境界オブジェクト

前章で述べた研究を複数の研究主体 (機関や団体) で協調的かつ競争的に進めていくために，我々はタグ付きコーパスと可視化プラットフォームを提供を行っている．これらは，異なる研究主体間の情報共有や相互理解を助け，円滑な課題遂行を促す境界オブジェクトとして機能することが期待される．以下にその詳細について述べる．

4·1 タグ付きコーパス

我々が配布しているタグ付きコーパスは，情報編纂に関わる研究コミュニティ内で対象分野 (領域) に対する一定の合意を形成する役割と同時に，様々な情報抽出処理の中間結果という役割をも担う．

配布しているタグ付きコーパス は 1998 年から 1999 年の 2 年間分の毎日新聞を元に，ガソリン価格やパソコン出荷状況など 27 トピックについて時系列になっている記事を収集し，各トピックにつき 3 つ前後の統計量を選んで，これらの統計量の可視化に必要な要素に対して人手でタグを付与したものである．一例として日経平均株価の記事 (毎日新聞 1999 年 5 月 29 日より) にタグが付与されたものを示す．

```
<unit stat="日経平均株価">
  <date gra="日" abs="19990528">28 日</date>
  の東京株式市場は,
  <del type="rsn">
    前日に米国のダウ工業株 30 種平均が今年最大の
    下げとなったのを受け,
  </del>
  <name>日経平均株価</name>は,
  <date gra="日" abs="19990331">
    3 月 31 日
  </date>以来
  <dur gra="月">2 カ月</dur>ぶりに終値で
  <val> 1 万 6000 円台</val>
  を割り込んだ
</unit>
```

この例では，<unit> タグ要素が日経平均株価であり，それに対して日付 (<date>タグ要素)，主題 (<name>タグ要素)，値 (<var>タグ要素) にタグが付与されている．このタグ付きコーパスは，MuST で最初に配布した版のタグ付きコーパスに一部拡張を施したものである (最初に配布した版のタグの詳細は文献 [加藤 04] を参照されたい)．拡張されたのは，<del> タグに属性が付与されたこ

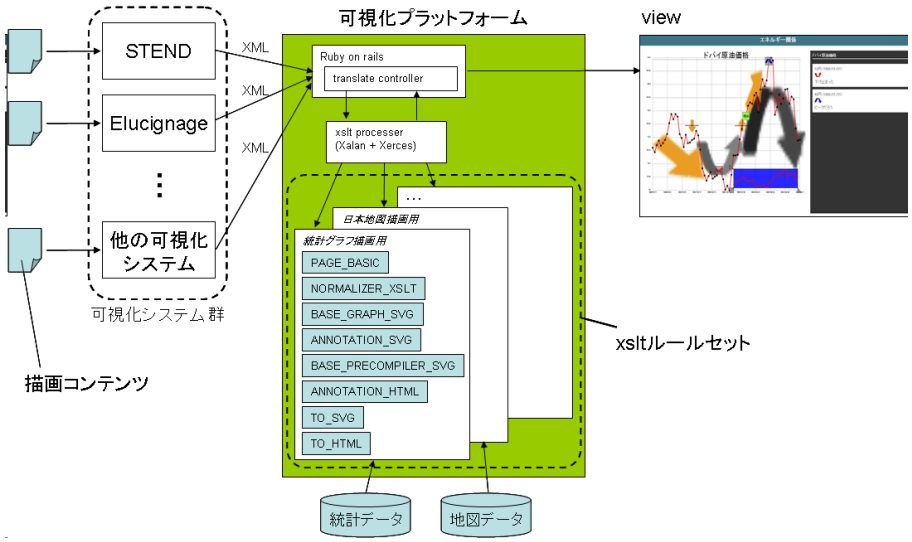


図 2 可視化プラットフォームの構成

とである．以前の仕様では <del>タグはグラフ描画に関係しない内容部分を特定する目的で導入されていた．このタグが付与された箇所は主に，なぜそのような事象が生じたかや，記述された内容の出典は何かなどが記述されていた．しかし，3 章 で述べたようにこれらはユーザの情報アクセスにおいて情報の内容を理解するうえで重要な手がかりとなるものである．この観点から，<del> タグ要素の属性として，rsn(理由「需要の伸びを受けて」「ダウ平均が下げたことにより」)，src(出典「通信白書によると」「総務省の調査で分かった」)，sit(状況「加入者増加で電話が掛かりにくくなっている」)，evl(判断「今後も下げ続ける見通し」)，oot(他項目「そのうちドコモのシェアは〜」)，rsn(その他) を付与し，視覚表現生成時に利用できるように拡張した．

テキストの可視化に係わる先行研究では，テキスト集合中の単語の出現頻度や共起関係に着目してテキスト同士の関連性を可視化するもの (例えば [Takama 03]) やユーザの興味や焦点に応じてテキストの易読性を向上させるもの (例えば [Suh 02]) など，内容の意味理解に踏み込まずに容易に獲得できる統計値やユーザによる指定に基づいたものが多い．それに対して，テキスト内容の意味理解に基づく可視化を行うものは，例えばタグ付き文書から概念図を生成する研究 [村山 99] などのように豊富な背景知識を用意し極めて限られた状況下での可視化表現生成が試みられているだけで，汎用的な可視化表現生成の枠組になり得る研究はまだ十分蓄積されていない．このコーパスの利用はこの課題の難しさを低減させ，意味理解に基づいた汎用的な可視化表現生成技術を成熟させていくうえで役立つと考えている．

4·2 可視化プラットフォーム

情報可視化システムの研究は，その特徴ゆえ視覚的に訴えることに重きが置かれる傾向にある．そのため，個別のシステムを取り上げると興味深く感じられるものが多い半面，複数のシステムを比べようとした場合，扱っているデータや機能が全く異なるため比較しにくいという問題があった．また，同じデータを扱う可視化システム同士を比較する場合であっても，見た目や操作性がその判断に影響する懸念があるため [Pillat 05]，統制された被験者実験のデザインが難しい．そこでこのような問題を解決するため，可視化システムの Look & Feel を共通化し同じ条件の下で可視化表現を比較できるように，異なるシステムがその上で動作する可視化プラットフォームを提供する．この可視化プラットフォームは統計情報の描画や拡大/縮小処理など，異なる可視化システムで共通に利用される機能をライブラリとして利用可能にすることで (1) 見た目や操作性に関する統一性を図り，異なるシステム同士の比較が容易になる，(2) 個々の研究者が各々の環境や独自のプラットフォームで実装した可視化システムを移植する際の負担が軽減される，というふたつの効果が期待される．

これまでも，情報可視化システムの要素技術の利用の容易化や簡便な可視化ツールの提供は様々な形で行われている．例えば，可視化ライブラリや API の提供 (project SIMILE[SIM 05] の timeline や timechart，Maryland 大の Piccolo toolkit[Bederson 04] など)，データを与えることでグラフを生成する WEB サービス (Open Flash Chart など)，可視化表現の生成に特化したプログラム言語の開発 (Design by Numbers[Maeda 99] や processing[Reas 07] など) がその一例である．これらの主眼は可視化表現生成の容易化を図る点にある．一方我々が提案する可視化プラットフォームの主眼は，評価のためのシステム比較の容易化と，情報抽出や要約技術などとの連携の容易化にある．

可視化プラットフォームの構成を図 2 に示す．この可視化プラットフォームは個々の可視化システムから XML 形

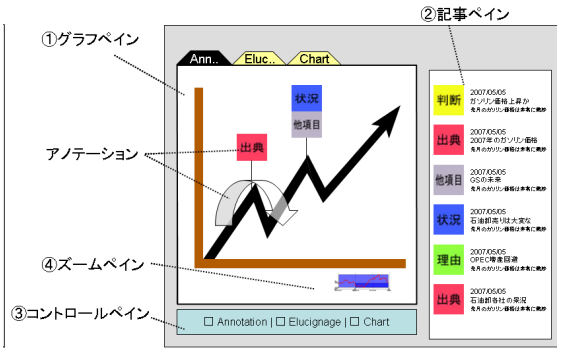


図 3 可視化プラットフォームによって生成される view

式の描画コンテンツを受け取り，可視化コンテンツ (view) を生成する機能を持つ．

各可視化システムは各々のシステムが規定する形式で描画コンテンツを生成するための情報 (テキストから抽出した時間表現や定性表現など，図 1 の「抽出されたデータ」に相当するもの) を受け取ると，XML 形式に変換し可視化プラットフォームに渡す．この XML には，可視化プラットフォームが提供する形式 (共通フォーマット) での記述に加え，SVG (Scalable Vector Graphics) フォーマットでの記述を埋め込むことが可能になっている．これにより，共通フォーマットを使うことによる処理の低減と生成する視覚表現の自由度とのバランスを図っている．すなわち，各可視化システムは必要に応じて可視化プラットフォームの枠を超えたより柔軟な可視化を行うことも可能である．

生成される view は図 3 のような構成をしている．図 3 中① のグラフペインが統計グラフやアノテーションが表示されるペインである．グラフペインは図 4 のようにレイヤ構造となっている．このレイヤのうち，統計情報レイヤと描画領域操作レイヤ (図 3 中④ のズームペインに相当) は可視化プラットフォームによって自動的に付与されるようになっている．その他のレイヤは各可視化システムによって規定されるため，任意のレイヤ数を持つことが可能である．図 4 では数値情報 (各可視化システムが独自に定義するもの) とアノテーションレイヤが設けられている．

図 3 中 ②の記事ペインには，関連する記事の一覧がスニペットを伴って表示される．スニペットはユーザに元記事を参照する価値があるかどうかの判断材料として提示されるものであるため，指示的 (indicative) な要約が表示されることを想定している．

図 3 中 ③ のコントロールペインは，レイヤの可視/不可視を操作するためのペインであり，可視化プラットフォームによって自動的に生成される．

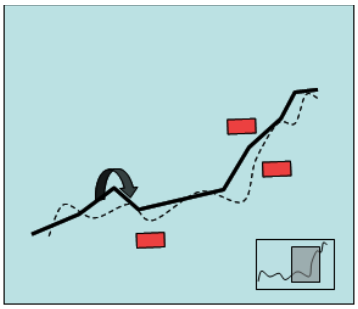
可視化プラットフォームの処理手順について簡単に説明する*3．まず，ユーザが XML フォーマットの可視化コンテンツ (各可視化システムの出力に相当) を ruby rails にアップロードする．rails 内にある translate コントローラがそれを受け取り，XSLT プロセッサ (Apache Xalan + Xerces を使用) に渡す．XSLT プロセッサは，XML ファイル中に記述されている可視化タイプ (e.g., “折れ線グラフ”, “日本地図”) から適用する XSLT モジュール群を判断し受け渡す．例えば，統計グラフ (折れ線グラフ) の描画に関する処理として，現時点での実装では 8 つの XSLT モジュールを用意している．これらのモジュールは，各々別個の可視化表現を行うものではなく，段階的に処理を行うために便宜的に分けているものである．各モジュールの簡単な XSLT 処理内容を以下に示す．

(1) PAGE_BASIC :
基本的なレイアウトを生成する．

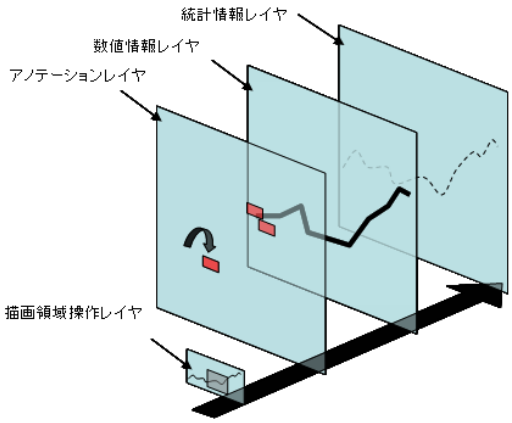
(2) NORMALIZER_XSLT :
データタイプが datetime に設定されている値 (“2007/11/09”) を描画平面上の絶対座標に変換する．

(3) BASE_GRAPH_SVG :
チャート部の生成 (折れ線グラフ描画，両軸設定，ラベル配置など) を行う

(4) ANNOTATION_SVG :
XML で記述されたアノテーションをチャート部



(a) グラフペイン



(b) グラフペインのレイヤ構造

図 4 グラフペインの構成

*3 現在は開発段階のため，単体でのみ機能する．そのため，XML を直接 rails にアップロードする形式になっているが，将来的には前段階の処理と連携し動作するように拡張する予定である．

上に配置する際の設定を行う．

(5) BASE_PRECOMPILER_SVG :
SVG 出力のための下準備 (SVG エレメントの定義など) を行う．

(6) ANNOTATION_HTML :
XML 内にあるアノテーションからアノテーション一覧に配置する際の設定を行う．

(7) TO_SVG :
SVG の出力を行う．

(8) TO_HTML :
HTML の出力を行う．

図 5 に動向情報を可視化するシステム STEND[松下 06] と Elucignage[松下 07] の出力画面を示す．STEND は文書情報 (新聞記事) から抽出した数値情報と定性情報とを表示する可視化システムであり，Elucignage は同じく時系列数値情報の上に，関連する言語情報をその内容に応じたアイコンの形式で提示する可視化システムである．これらの可視化システムは同一の統計情報を対象とし，その上に各々付加情報を重畳して view を生成している．STEND は JAVA の描画ライブラリを，Elucignage は action script (ver. 2.0) を各々用いて view を生成している．図 5 からわかるように，見た目の印象は大きく異なる．これらのシステムの view を本稿で提案する可視化プラットフォームを用いて生成すると，図 6 のようになる．これによって，二つのシステムの比較が見た目に左右されずより公平なかたちで行えるようになる．なお，現状では棒グラフや折れ線グラフなど統計グラフを対象とした可視化を提供するが，将来的には地図などの地理情報の表現およびそれらの連携を考えている．

可視化システムの評価においては，(1) 可視化の要素技術を取り出して比較する，(2) ユーザ観察によりシステムの問題点を洗い出す，(3) 統制実験により二つ以上のシステムを比較する，(4) 実際の場面で利用して有用性を検証する，などの方法が採られている [Plaisant 04]．本稿で提案した可視化プラットフォームはこのうち (3) の評価方法を，個々の研究者で閉じた形ではなく，類似した興味を持つ複数の研究者が関与しコミュニティとして実施できるようにすることを意図したものである．もちろん，可視化の Lool & Feel がインタラクティブな知識発見の場面では大きな影響を与えることは十分に考えられるため，可視化プラットフォームを前提とすると，実際の利用場面での有用性低下を招く懸念が指摘されよう．そのため，我々のスタンスとしては，各研究者や研究機関に可視化プラットフォームの使用を強いるのではなく，あくまでもコミュニティ内での客観的評価の基盤としての利用を推奨するという立場である．提案する可視化プラットフォームが XML 形式の可視化コンテンツを入力として考えているのは，個々の研究者が各々の環境や独自のプラットフォームで可視化システムを実装した場合であっても，その出力を XML 形式で出力できる機能を

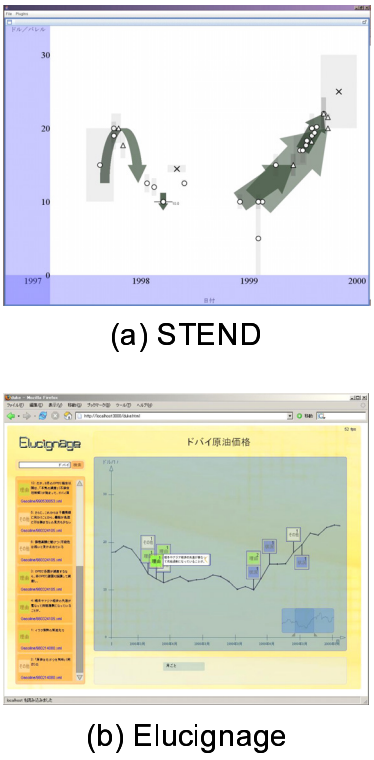


図 5 従来の STEND と Elucignage の出力例

付与しておけば，この可視化プラットフォームを比較的容易に利用できると考えているためである．この点に関しては，今後実際に各研究者や研究機関に利用してもらうことで検証していきたいと考えている．

5. 評価なきワークショップから評価ワークショップへ

コミュニティの具体的形態のひとつにワークショップがある．ワークショップでは，コミュニティのメンバが参照モデルと境界オブジェクトを共有し，互いに影響を与えながら研究を進めることが期待される．研究を遂行する場面では，その段階やメンバ間の関わり方によって様々なインタラクションが考えられるが，それに応じてワークショップの位置づけも変化する．

特定の研究課題について協調的かつ競争的に研究を進めていくための枠組みとしてのワークショップとしては，TREC, NTCIR のような評価ワークショップが有名である．その意義は以下のようにまとめられる [神門 02].

- 研究者フォーラム，“Who is who” 環境の創生
- 研究データの蓄積
- 研究の動機付け，特定研究課題への集中的研究
- 研究モデルの提示

評価ワークショップにおいては，大規模なテストコレクションの構築とそれに基づいた客観的な評価を行うことで，研究データの蓄積と研究の動機付けが実現される．このことは評価型ワークショップの大きな利点であるが，

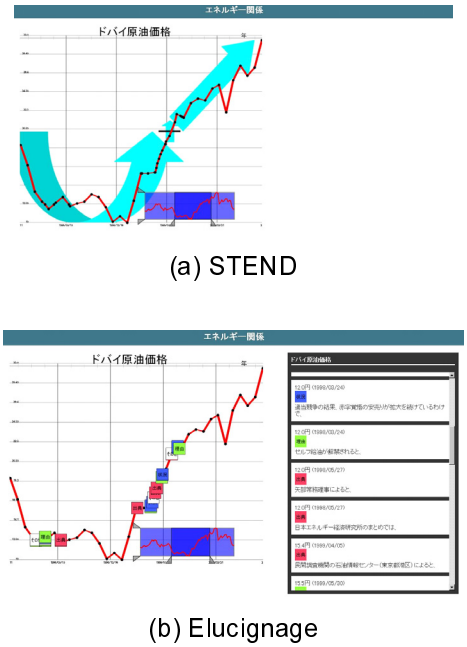


図 6 可視化プラットフォームを用いた場合の STEND と Elucignage の出力例

常にそのような評価が行えるわけではない．特にその研究領域が複数の既存分野をまたがる場合は共有された評価基準が存在せず，要素技術の研究が主となるため，客観的共通的な数値評価が難しい．しかし，このようないわゆる黎明期の研究領域でも，ワークショップはコミュニティの形成，研究領域の認知度の向上等に有益である．そのワークショップは必ずしも評価ワークショップである必要はないが，評価に代わる求心力をどこに求めるかが問題となる．

著者らは，情報編纂技術へと繋がる「MuST：動向情報の要約と可視化に関するワークショップ」を評価なきワークショップとして提案した．そして，その求心力を，参照モデルとして全体システムの構想と枠組みを提示すること，それに基づいて，研究の素材となる入力データにとどまらず，提示した枠組みにおいてハブとなるような，つまり，あるコンポーネントの出力であり別のコンポーネントの入力となるような処理の中間結果のコーパスの共有に求めた．その後の可視化プラットフォームの提供も同じ流れの中にある．

そのような評価なきワークショップを通じて，共通の素材を用いて，緩い意味で共通の課題に取り組むことを通じて，何人かの研究者 (研究組織) に共通する研究課題が生まれる．それらについて，評価ワークショップを行っていくという展開を考えている．

自然言語処理や機械学習の分野では shared task という形である意味で絞り込まれたタスクについて，様々な取り組みをすることで資源の共有や手法の進展を図り，形成されたコミュニティで得られた知見や方法論を総括することで更なる発展へと繋げる試みがある (e.g., [Ireson 05], [Belz 06]) が，それ以前，つまり，何を task として share するかを検討することをコミュニティ内で議論する場は多くない．情報編纂コミュニティのように異なる研究分野や専門に跨る境界領域研究では，shared task を事前に決めるのではなく，コミュニティの発展に伴って共通する関心や興味が明確になり，そこからコミュニティ内の議論を通じて shared task が規定されていくという方向性が望ましいのではないかと考えている．

実際，動向情報を対象とした情報抽出においては，テキストからの時系列統計情報の抽出，数値情報列の文章化，テキスト情報と数値情報のアライメントが基本技術として複数の研究者に取り上げられた．この課題について，評価基準を検討し，より一般性のある研究課題として洗練することを目的に評価ワークショップを展開している．この評価ワークショップは NTCIR-7 のタスクとして行われている．

6. これまでの成果と今後の方向性

本章では，まず情報編纂の先駆けとなる「動向情報の要約と可視化ワークショップ」の観点からの総括を述べ，次に近未来チャレンジとしての総括とこれからの展望について述べる．

6·1 MuST ワークショップ

MuST ワークショップでは様々な研究が進められている．ワークショップの開始時点で考えていた言語情報と非言語情報を総合的に処理し活用するというマルチモーダル要約の範囲を超えて，動向を見つけ出し分析するというトレンド・マイニングにまで広がりを見せている．このような広がり自体が，情報編纂への展開を動機づけたと言ってよい．これら様々な研究は既存の研究分野としては多岐にわたるが，それらに携わる研究者が動向情報という共通の関心で集まり，議論する場を創造できたという点で MuST の果たした役割は大きい．そこに集まった研究者は，MuST データセットに含まれている新聞記事集合を眺めることで，何が動向情報であるか，何が動向情報に係わるかの認識を共有しており，それが議論を活性化していると考えるが，このような認識を作り出したという点は MuST データセット (4.1 節で述べたタグ付きコーパスとそれに関連する数値データ) の功績と考える．

一方で，データあるいはその仕様として MuST データセットの注釈が期待通りの役割を果たしたかには疑問は残る．MuST データセットの注釈はテキスト要約もしくは情報抽出の中間データという位置づけであったが，情報抽出の研究においては，データセットの注釈ではなく汎用の固有表現抽出システム等による結果を用いるものが多く，注釈結果を前提にその後の処理を研究対象するもの ([太田 07] はその例である) は少ない．ただし，選

ばれたトピックや新聞記事のセットについては利用が多く，共通の入力，処理対象としての役割は果たしたと考えられる．もちろん，新聞記事セットについても多量であることが重要な場合は他の資源が求められているが，MuST データセットの共通の処理対象としての役割は，動向情報についての共有認識を作り出したことに次ぐ大きな役割であろう．加えて，情報可視化に重点をおく研究では XML パーザを用いて MuST データセットから可視化データの取り出しが行われているし，更に，研究における初期の分析，評価のための参照，正解データの作成としては頻繁かつ有効に利用されているという印象を受ける．その点で注釈付けもその役割を果たしていると判断できる．

ワークショップのもうひとつの目的であるツールやデータの共有については，まだ充分とはいえない．多数のデータの提供はあるもののその再利用は残念なことに殆どない．２回のサイクルで個々の研究の関心が明らかになりその内容も充実したことで，提供できるデータやツールの質も向上してきていると思われるので，今後の活用に期待していきたい．また，これからこのテーマに関心を持つであろう新規参入の研究者達による利用も期待したい．

6·2 近未来チャレンジ

筆者らが「情報編纂の基盤技術」を近未来チャレンジのテーマとして提案してから，本稿執筆時点で 2 年間が経過している．5 年間という近未来チャレンジのスコープからすると，折り返し地点といえる段階である．ここで，これまでの状況を総括し今後の展開・展望について俯瞰したい．

1 章で述べた目標達成のためのステップのうち，Step 1. に関しては，MuST「動向情報の要約と可視化に関するワークショップ」での取り組みも含め，(1) 言語情報からの動向情報要素の獲得手法，(2) 数値データに基づく言語表現の生成手法，(3) 動向情報の可視化手法，など情報編纂で核となり得る要素技術が提案され，その精緻化が進められている．特に (1) と (2) に関しては，5 章でも述べたように shared task として評価ワークショップを行える段階に達している．また (3) に関しても，(1) や (2) の技術との連携を視野に入れた試みが進められている (例えば [松下 06])．これらから，Step 1. に関しては所期の計画に沿って進捗していると言えよう．今後は，評価ワークショップを通じて得られる知見に基づき基盤技術として整理していくとともに，利用可能なツールやデータセットの整備を図りたいと考えている．

Step 2. に関しては，技術文書を対象としてそこから製品や技術の好ましい性質や機能に関する言及を抽出する技術 [西山 08] や，会議の場面で見出されたアイデアや問題点などをドキュメント生成などに再利用するための技術 [土田 07] など，個々の技術としては興味深い提案がされているものの，基盤技術の確立には至っていない．一方で，NTCIR-6 のパイロットタスクとして opinion analysis が提案され，NTCIR-7 で正式なタスクとして多言語の opinion analysis が行われるなど，意見や評判などをテキスト中から取り出す研究に注目が集まっている．これらの研究の進展を鑑みつつ，動向に関連する情報の抽出・可視化のための基盤技術の確立を目指したいと考えている．また，これらを扱う研究では，その情報の産出者や利用者によるボトムアップなアノテーション付与やそれに基づく組織化など，ユーザによって与えられるメタ情報の利用を狙ったものもある ([土田 07] などはその例である)．このような，情報編纂の協創的側面に関しては提案の段階では想定していなかったが，昨今注目を集めている CGM (Customer Generated Media) に代表されるように，情報の利用者がコンテンツ生成に積極的に関与することが当り前になりつつある現状では，考慮に含めて然るべきであろう．具体的には，どのようにユーザの参加を促すか (インセンティブ設計)，ユーザによって生成・付与された情報をどのように活用するか (メタ情報の管理) などが挙げられる．

Step 3. に関しては，工藤らによって高精細な仮想空間上で大量の画像データを閲覧するために，それを閲覧するユーザのアクセス軌跡を画像データ鑑賞の「文脈」として利用する試み [工藤 08] が提案されているに留まっている．2 章 で述べたように，画像データなどのパタン情報からシンボリックな意味を抽出する研究は現時点では直接の関心ではない．しかし工藤らの問題設定のように，現実にはアノテーションが全く付与されていない非言語情報が大量に存在しており，近未来チャレンジの評価指標のひとつである「社会貢献」を鑑みると，情報編纂の対象としてそれらの情報を除外することはできない．この場合も，Step 2. 同様，情報利用者の情報編纂への積極的な関与を考慮に入れることがその課題への切口となり得よう．ただし，個々のパタン情報への明示的なアノテーション付与といった直接的な方法は，一定規模の情報利用者の関与が前提となるため，近未来チャレンジの意図する時間的スパンでは難しいと考える．非言語情報に関しては，工藤らのアプローチのように大量情報に対するユーザの振るまい・体験を通じて，情報編纂に活用できるメタ情報の蓄積・再利用のための方法論についても検討すべきであろう．

6·3 情報編纂技術の今後の方向性

6·1 節および 6·2 節を踏まえ，情報編纂の全体の構成および今後のマイルストンを再考する．

図 7 は，文献 [加藤 06] で示した情報編纂技術の構成を詳細化し，ユーザの関与をスコープに加えたものである．したがって，コミュニティ支援技術が新しく追加されたことになる．二重線で囲まれている箇所が技術項目であり，各技術項目の中の要素技術 (図中，角の丸い四角形) のうち，太線で描かれているものは特に今後の発

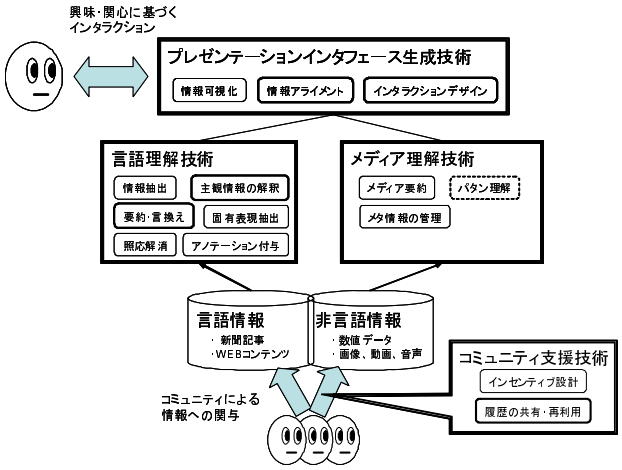


図 7 情報編纂技術の構成

展を期待するもの，点線は関連する要素技術ではあるが，当面の関心から除くものを各々意味している．

このような全体構成の下，節目となる 3 年後 (近未来チャレンジへ提案してから 5 年後) に向けて，以下の 2 点を目標とする．

(1) 情報編纂のひとつの具体化としてのシステムの実現
- 数値情報とテキスト (事実，意見，予測，変化の要約・評価など) を組み合わせ，ユーザの関心や意図に応じた情報を提示する
- 変化に関する言語表現を視覚的な表現に返還する
- ユーザの興味や関心の変化に応じてインタラクティブに提示情報を切り替える

(2) 情報編纂の様々な可能性の洗い出し
- 研究コミュニティで提案された技術やシステムイメージをきっかけにコミュニティ内での議論を通じて新たな可能性を洗い出す
- コミュニティ内の共通の関心や研究課題を整理し，その競争的協調的連携の一助となる評価指標を提案する

そして，これらの取り組みを通じて，より広範な非言語情報の利用へと展開していきたいと考えている．

7. むすび

本稿では，情報編纂技術を協力的かつ競争的に行うために，我々が現在進めている試みについて述べた．情報編纂技術に関する研究を進める上で，非言語情報と言語情報の協調によるモード横断性と，情報の理解・概観の支援と詳細な情報へのアクセス支援とをシームレスに繋ぐ情報活用支援の連続性とに力点を置き，それを進める上での方法論として，(1) 参照モデルの設定，(2) 境界オブジェクトの提供，(3) 評価型ワークショップの展開を行った．この取り組みを通して，情報編纂技術が進展することを期待している．

謝辞

本研究を進めるにあたって，「MuST: 動向情報の要約と可視化に関するワークショップ」参加者の皆様，ならびに近未来チャレンジ「情報編纂の基盤技術」参加者の皆様と活発な議論の場を持つことができました．その際に頂いた様々な御示唆に感謝します．また，有益なご指摘を賜りました査読者の皆様に感謝します．本研究の一部は国立情報学研究所の公募型共同研究によって支援されています．ご支援を感謝いたします．

◇ 参 考 文 献 ◇

[Arias 00] Arias, E., Eden, H., Fischer, G., Gorman, A., and Scharff, E.: Transcending the Individual Human Mind — Creating Shared Understanding through Collaborative Design, *ACM Trans. on Computer-Human Interaction*, Vol. 7, No. 1, pp. 84–113 (2000)

[Bates 89] Bates, M. J.: The Design of Browsing and Berrypicking Techniques for the Online Search Interface, *Online Review*, Vol. 13, No. 5, pp. 407–424 (1989)

[Bederson 04] Bederson, B. B., Grosjean, J., and Meyer, J.: Toolkit Design for Interactive Structured Graphics, *IEEE Trans. on Software Engineering*, Vol. 30, No. 8, pp. 535–546 (2004)

[Belz 06] Belz, A. and Kilgarriff, A.: Shared-task evaluations in HLT: Lesons for NLG, in *Proc. INLG'06*, pp. 133–135 (2006)

[Card 99] Card, S. K., Mackinlay, J. D., and Shneiderman, B. eds.: *Readings in Information Visualization — Using Vision To Think —*, Morgan Kaufmann Publishers (1999)

[Fayyad 96] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. eds.: *Advances in Knowledge Discovery and Data Mining*, The MIT Press (1996)

[Fischer 01] Fischer, G.: External and Shareable Artifacts as Opportunities for Social Creativity in Communities of Interest, in *Proc. 5th Int. Conf. on Computational and Cognitive Models of Creative Design*, pp. 67–89 (2001)

[藤本 05] 藤本 和則, 本村 陽一, 松下 光範, 庄司 裕子：意思決定支援とネットビジネス, オーム社 (2005)

[Ireson 05] Ireson, N., Cirevegna, F., Califf, M. E., Freitag, D., Kushmerick, N., and Lavelli, A.: Evaluating Machine Learning for Information Extraction, in *Proc. ICML2005*, pp. 345–352 (2005)

[神門 02] 神門 典子：NTCIR とその背景–情報アクセス技術の評価ワークショップとテストコレクション, 人工知能学会誌, Vol. 17, No. 3, pp. 296–300 (2002)

[加藤 04] 加藤 恒昭, 松下 光範, 平尾 努：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, 信学技報 NLC2004-25, pp. 13–18 (2004)

[加藤 06] 加藤 恒昭, 松下 光範：情報編纂 (Information Compilation) の基盤技術, 2006 年度人工知能学会全国大会（第 20 回）論文集, 1D3–02 (2006)

[加藤 07a] 加藤 恒昭, 松下 光範：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, ヒューマンインタフェース学会誌, Vol. 9, No. 4, pp. 65–70 (2007)

[加藤 07b] 加藤 恒昭, 松下 光範：変化を基本単位とした時系列情報の抽出と可視化, 動向情報の要約と可視化に関するワークショップ第 2 回成果進捗報告会予稿集, pp. 33–38 (2007)

[Kato 07c] Kato, T., Matsushita, M., and Kando, N.: Fostering Multi-Modal Summarization for Trend Information, in *Proc. KES2007, Part 2*, pp. 377–386 (2007)

[工藤 08] 工藤 佑太, 川嶋 稔夫, 柳 英克, 中島 秀之：高精細仮想空間におけるアクセス軌跡共有による情報の編纂, 2008 年度人工知能学会全国大会（第 22 回）論文集, 3K3–08 (2008)

[Maeda 99] Maeda, J.: *Design by Numbers*, MIT Press (1999)

[Marshall 05] Marshall, C. C. and Bly, S.: Saving and Using Encountered Information: Implications for Electronic Periodicals, in *Proc. CHI2005*, pp. 111–120 (2005)

[松下 06] 松下 光範, 加藤 恒昭：数値情報の補填とグラフ概形の示唆による複数文書からの統計グラフ生成, 知能と情報, Vol. 18,

No. 5, pp. 721–734 (2006)
[松下 07] 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, 2007 年度人工知能学会全国大会, 3H8–3 (2007)
[村山 99] 村山 正司, 中村 裕一, 大田 友一：概念図の自動生成によるタグ付文書の可視化, 電子情報通信学会技術報告 TL99-25, pp. 51–58 (1999)
[西山 08] 西山 莉紗, 竹内 広宜, 渡辺 日出雄, 那須川 哲哉, 武田 浩一：技術文書マイニングのための特長表現抽出, 2008 年度人工知能学会全国大会（第 22 回）論文集, 3K3–02 (2008)
[太田 07] 太田 彰, 福本 淳一：文書中の数値的特徴を用いた情報可視化, 動向情報の要約と可視化に関するワークショップ 第 2 回成果進捗報告会予稿集, pp. 13–16 (2007)
[奥村 05] 奥村 学, 難波 英嗣：テキスト自動要約, オーム社 (2005)
[Pillat 05] Pillat, R. M., Valiati, E. R. A., and Fraitas, C. M. D. S.: Experimental Study on Evaluation of Multidimensional Information Visualization Techniques, in *Proc. CLIHC2005*, pp. 20–30 (2005)
[Plaisant 04] Plaisant, C.: The Challenge of Information Visualization Evaluation, in *Proc. AVI2004*, pp. 109–116 (2004)
[Reas 07] Reas, C. and Fry, B.: *Processing: A Programming Hnadbook for Visual Designers and Artists*, MIT Press (2007)
[Shneiderman 98] Shneiderman, B.: *Designing the User Interface*, Addison-Wesley, third edition (1998)
[SIM 05] Project SIMILE: Semantic Interoperability of Metadata and Information in unLike Environments, MIT SIMILE Proposal (2005)
[Suh 02] Suh, B., Woodruff, A., Rosenholtz, R., and Glass, A.: Popout Prism: Adding Perceptual Principles to Overview+Detail Document Interfaces, in *Proc. CHI2002*, pp. 251–258 (2002)
[Takama 03] Takama, Y. and Hori, T.: Application of Immune Network Metaphor to Keyword Map-based Topic Stream Visualization, in *Proc. CIRA2003*, pp. 770–775 (2003)
[土田 07] 土田 貴裕, 友部 博教, 大平 茂輝, 長尾 確：会議コンテンツの効率的な再利用に基づく知識活動支援システム, 2007 年度人工知能学会全国大会（第 21 回）論文集, 2H5–01 (2007)

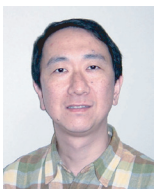
〔担当委員：阿部 明典〕

2007 年 12 月 18 日　受理

著　者　紹　介

松下　光範（正会員）

1969 年生まれ．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) 入社．2008 年より関西大学総合情報学部准教授，現在にいたる．インタラクションデザイン，自然言語理解に関する研究に従事．博士 (工学)．1996 年度人工知能学会全国大会優秀論文賞，2001 年度人工知能学会全国大会ベストプレゼンテーション賞，平成 14 年度情報処理学会論文賞ほか各受賞．情報処理学会，日本ファジィ学会，日本バーチャルリアリティ学会，ACM 各会員．



加藤　恒昭（正会員）

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助教授，現在に至る．自然言語理解，対話処理，マルチモーダルコミュニケーションに関する研究に従事．工学博士．電子情報通信学会，人工知能学会，言語処理学会，ACL 各会員．