

特集 「研究会総覧」

情報編纂研究会 (SIG-IC)

Special Interest Group on Information Compilation

加藤 恒昭
Tsuneaki Kato

東京大学, 情報編纂研究会主査
The University of Tokyo.
kato@boz.c.u-tokyo.ac.jp

情報編纂研究会 (SIG-IC) は, 2009 年度発足, 11 月にキックオフシンポジウムを開催したばかりのできたてホヤホヤの研究会である。ただ, この研究会は, 人工知能学会全国大会近未来チャレンジで, 2006 年に提案を行い [加藤 06], 現在 3 回 (提案時を含め 4 回) のサイババルを続けている「情報編纂の基盤技術」と密接に関連しているのだから, これを含めると, その歴史はもう 4 年に近いことになる。さらに, その関心は, 2004 年 11 月に情報処理学会自然言語処理研究会・電子情報通信学会言語理解とコミュニケーション研究会の合同研究会で提案し [加藤 04], 国立情報学研究所が主催する NTCIR (情報アクセス技術の評価に関するワークショップ) のパイロットタスクとして, その後正式タスクとして実施してきた MuST (動向情報の要約と可視化に関するワークショップ) [Kato 07, Kato 08, 加藤 09] と継承発展の関係にある (MuST は 2008 年 12 月の NTCIR-7 ワークショップミーティングをもって, 一時休息中)。そこから考えれば, 5 年以上の活動を経ての満を持しての発足ともいってよい。

「情報編纂」という語は, 近未来チャレンジ提案時に私達 (松下光範氏 (当時 NTT, 現在関西大学) と本稿著者) が考えだした (つもりでいる) もので, 実は「情報編集」にしようかとも悩んだ末, エディタやワープロ関連の技術と誤解されても困るところらを採用したという経緯がある。そのことから察していただけるように, 情報編纂は「さまざまな情報を編集・編纂して, わかりやすくしよう」という技術である。なぜわかりやすいことが大事かといえば, 必要な情報か否かを少ない負荷で判断でき, その内容を容易に把握し理解できることにつながるからである。そして, 編集・編纂したものをどうやってわかりやすく示すかという視覚的な手段を駆使してということになる。さらに, 編集・編纂の対象となる情報はさまざまということで, 研究会ホームページ^{*1}にもある設立の趣旨での以下のような説明となる。

情報編纂は, 対話的探索的な情報アクセスにおける言語情報と非言語情報の有機的な利用に関する技術で, 以下の三つの特徴をもつ。

- 言語情報と非言語情報をシームレスに扱う

- 情報アクセスのインタフェースとして, 情報可視化, マルチモーダル要約, マルチメディアプレゼンテーションなどを通じて, 言語情報, 非言語情報を有機的に利用する

- 利用者の情報アクセス過程に追従して, もしくはそれをリードして, 概要把握から詳細情報の取得までを一貫して対話的に支援する

キックオフシンポジウムでもご指摘いただいたのだが, 「編纂」という語は, 体系的にまとめあげる, 一冊の本にする, 一貫したストーリーを付与することなどを連想させる。それでは, 情報アクセスというプロセスの中で, 情報要求をもった個人 (利用者) に適した形という動的で適応的なイメージが浮かび上がってこない。対話的・探索的な環境でのわかりやすさという点は MuST を運営している過程で強調されてきたもので, 近未来チャレンジ提案時にはずいぶんと前面に出されていたのであるが, 「編纂」という語にはそれが映し出されていないのが残念である。とはいえ, この一点を除けば, 編纂 (compilation) という概念は, 雑多な情報を一つの観点でまとめあげ, わかりやすく見せる, 概観したり全体像を理解することを助けるという印象を正確に伝えているように思う。

今回, 人工知能学会の研究会, 特に第三種研究会として活動を開始したわけであるが, その目論見は, 以下の 4 点にある。

- 情報編纂技術の入門書 (既存技術との関係や差異の明確化や, 事例の紹介を通じて技術のコンセプトを明らかにし, 多くの研究者にその分野に取り組む動機づけを与える方向性をもつもの) の執筆
- 情報編纂技術の展開を支援するインフラ (ツール群, プラットフォーム, コーパス) の構築
- 情報編纂技術の評価についての枠組みの提案
- 情報編纂技術をアピールするパイロットシステムの構築 (テーマ研究の実施)

MuST にしても近未来チャレンジ「情報編纂の基盤技術」にしても, 私達がやってきているのは, いわば音頭取りであって, コンセプトや研究素材を提案・提供しているだけである。そして, そのコンセプトに共感し, 研究素材に興味をもってくださった方々が, 共通点はあるがさまざまなテーマで研究を展開してくださっている。

*1 <http://taurus.c.u-tokyo.ac.jp/SigIC/index.html>

そこで議論が始まり、そこにコミュニティが生まれる。そして、その中で私達自身も参加者として研究を進めている。

例えば MuST であれば、動向ということ、統計情報（製品のシェアや支持率）やイベント情報（地震の発生や台風の上陸）に関連した文書を扱うことが主だったので、以下のような研究が展開された。

- テキスト情報の要約、そこからの情報抽出として、文書からの特定の統計量に関する一連の情報抽出
- 時系列数値情報からそれを説明する文章を生成するという数値情報の言語情報化
- 情報を概観するためのプレゼンテーションの生成もしくはそれを参照して情報にアクセスするインタフェースの実現として、時空間にまたがる情報の対話的な可視化や時系列情報と言語情報の関連付け
- 動向の発見や統計量間の関係の抽出や文書集合に含まれる情報全体の概観など、動向マイニングと呼べるテキストマイニング

近未来チャレンジでは、さらに対象と関心を広げて、次のようなテーマが加わっている。

- 技術文書からその特徴を抽出し未来予測を行うテキストマイニング
- 議事録におけるトピックの推移や議論の文脈を可視化するなどの文書可視化
- 高精細仮想空間による図書館メタファを用いた情報アクセス環境の構築

このような研究形態の発展は素晴らしいことである。その一方で、このあたりで、これらの核となるものが何かを明らかにし、技術として展開していくための基盤を整理・整備することも必要だという思いがあり、研究会の発足となった。目論見の4点はこのことにつながっていて、情報編纂のコンセプトを明らかにし、さまざまな研究・技術を整理して情報編纂という視点での融合を図り、そのことの副作用として教科書が執筆される。その研究を進めるためのインフラであるツール群、プラットフォーム、コーパスについて、既存のものを整理するとともに必要なものを構築する。技術を評価する枠組みを定義する。そして、「これぞ、情報編纂！」というシステムを例示することで応用と展開の方向を示そうというものである（図1）。どれ一つとっても、十分に大きい活動テーマであり、例えば、評価の枠組みは、対話的なシステムの使い勝手という主観的な判断を含むものの定量的評価ということで、これまでさまざまな試みがなされていながら、なかなか決定打が出てきていない難し

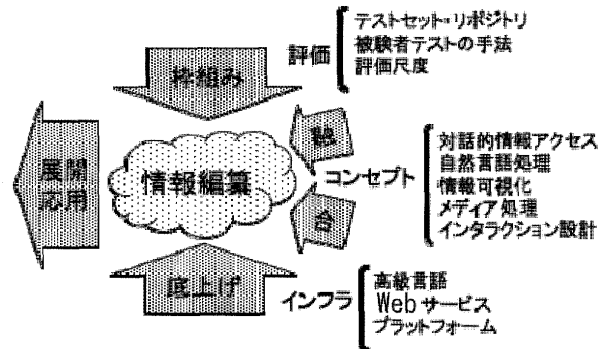


図1 情報編纂研究会の枠組み

い課題である。欲張りすぎであるのご批判もあるだろう中、うまく資源を配分しながら進めていきたいと考えている。

具体的な進め方としては、情報編纂につながる研究分野に関するチュートリアル、情報編纂に関わる研究報告を中心として、研究会のシンポジウムを開催して多くの研究者に情報発信するとともに、コアメンバを中心に評価に関する研究やインフラの構築を進めつつ、その報告やそれへのご意見をいただく議論の場としてもそのシンポジウムを活用していきたいと考えている。多くの人達の研究活動と有機的に関連する研究会とするとともに、NTCIRなどの研究活動とも連携していければとも考えている。

ご理解いただいたとおり「猪突猛進」, 「大言壮語」的な取組みではある。そうではあるが、そしてそれゆえに皆様の積極的なご支援をよろしくお願ひしたい。

◇ 参 考 文 献 ◇

[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, 情処学会自然言語処理研究会, 2004-NL-164, pp. 89-94 (2004)

[加藤 06] 加藤恒昭, 松下光範: 情報編纂 (Information compilation) の基盤技術, 第20回人工知能学会全国大会, 1D3-2 (2006)

[Kato 07] Kato, T., Matsushita, M. and Kando, N.: Expansion of multimodal summarization for trend information — Report on the first and second cycles of the MuST Workshop —, Proc. 6th NTCIR Workshop Meeting, pp. 235-242 (2007)

[Kato 08] Kato, T. and Matsushita, M.: Overview of MuST at the NTCIR-7 Workshop — Challenges to multi-modal summarization for trend information —, Proc. 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies, pp. 475-488 (2008)

[加藤 09] 加藤恒昭, 松下光範, 神門典子: 動向情報の要約と可視化とその展開 — MuST (動向情報の要約と可視化に関するワークショップ) 活動報告一, 情処学会自然言語処理研究会, 2009-NL-190-12 (2009)