

[IFAT]

[情報基礎とアクセス技術

研究会]

1. 最近 10 年間の動向

情報基礎とアクセス技術研究会 (IFAT) は、情報に関する原理を解明し、それに基づいて情報アクセスに関する技術の発展を目的としている。もともとは「情報学基礎 (Fundamental Infology: FI)」という名称の研究会であったが、情報検索や情報推薦をはじめとする情報アクセス分野の研究などを取り込む学際横断的な研究会を目指し、2010 年に現在の名称に変更された。また、ほぼ 同時期に「情報アクセスシンポジウム (IAS)」という IFAT 単独主催のシンポジウムが開始された。定例の研究発表会と同様の口頭発表で構成されたセッションに加えて、学生や若手研究者の教育を指向したチュートリアル、メンタリング (投稿前の原稿に対する指導)、英語限定セッション等の野心的なアイデアが投入された。

IFAT の状況を数字で確認する。表 1 は 2010~2018 年までの IFAT 登録者数、定例発表会と情報アクセスシンポジウム (IAS) における発表件数の推移である。シンポジウム開催などの新しい取り組みもあり、発表も一定数ある一方で、登録者数は概ね漸減している。

表 1: 登録、定例発表、IAS シンポジウム発表の件数 (斜線は IAS の開催なし)

年	2010	2011	2012	2013	2014	2015	2016	2017	2018
登録数	191	186	176	187	182	164	153	148	135
定例発表数	45	49	36	56	44	36	52	64	59
IAS 発表数	9	8	11	8	5				

2. 研究分野の変遷

IFAT に名称変更してからの 10 年間の研究発表がど

のように変遷していったのかを把握するため、2010~2019 年に IFAT で発表された 435 件の論文表題 (英語表題を除く) を以下に述べる手法で分析した。

「どのようなデータや要素技術を用いて研究を行っているのか?」は、その研究の重要な特徴のひとつであるが、このような情報は、論文表題に明示されることがある。例えば「新聞記事データを用いたテキスト平易化」という論文表題の場合、「を用いた」の直前にある「新聞記事データ」は「テキスト平易化」を実現するために用いられるデータであることを示す。筆者は、「を用いた」のような手がかり語に着目し、機械学習を用いた論文表題の構造を解析するシステムを構築している¹⁾。このシステムを用いて 435 論文の表題を解析し、論文中で使用されているデータまたは要素技術を自動抽出した。結果の一部を表 2 に示す。なお、抽出結果がデータであるか技術であるかは、筆者が判断した。

表 2 において、要素技術を見ると、2016 年以降に深層学習関連の用語が数多く出現していることがわかる。自然言語処理や情報推薦分野は、画像処理分野に比べ、少し遅れて深層学習のブームが来ているという話を聞いたことがあるが、今回の分析で改めてそのことが確認できた。この他、メタ情報を記述するためのデータ形式の一つである RDF (Resource Description Framework) が時々顔を覗かせるあたりは、「情報基礎」を冠する研究会らしい。

次に、データに着目する。Twitter や Wikipedia は 2010 年と 2019 年の両方に出現しているが、これらのデータを使った研究は昔からよく見かける。一方で、クエリログやアクセスログは 2010 年代前半に、コスメアイテムやファッション系データは 2010 年代後半に出現しているが、これは国立情報学研究所の情報学研究データリポジトリ (IDR) から 2017 年に公開されたホットペッパービューティーデータと関係すると思われる。

表 2: IFAT 研究会で発表された 435 論文の表題解析
結果

	データ	要素技術
2010	Twitter, Wikipedia, Q&A サイト	トピックモデル
2011	Twitter, Wikipedia, クエリログ, 電子書籍	
2012	Wikipedia, ソーシャルブックマーク, アクセスログ	RDF, クローラ, 統計的フレーズ翻訳モデル
2013	マイクロブログ, 位置情報付きテキスト, ユーザプロフィール	RDF, センサネットワーク
2014	Wikipedia, クエリログ, EC/観光サイト	トピックモデル, センサネットワーク
2015	新聞, Q&A サイト	RDF
2016	地理情報, ライフログ	多層パーセプトロン, GPU
2017	ジオタグツイート, ユーザ行動データ, コスメアイテム	深層学習, 機械学習, 視線計測, 感情分析
2018	ジオタグツイート, 感情辞書, 雑談内容, 日本語学習者会話	CRF, URL エンベディング, 類似画像抽出
2019	ファッション SNS, Twitter, Wikipedia	ニューラルネットワーク, LDA モデル, 機械学習

名称変更前の FI の時代は, 論文や特許や新聞などフォーマルなデータを扱った研究が多かったような印象がある。これに対し, 近年は SNS などのカジュアルなデータを扱った研究発表が増えている。これは, 最先端技術を指向した研究と開発を象徴している「(情報)アクセス」が研究会名に入ることにより, 以前よりも発表しやすくなったと感じる人が増えた

のかもしれない。過去に IFAT で発表したことのない新規の研究グループの論文も多く, 研究会としての間口が広がったことが実感できて嬉しく思う。

3. 今後の展望

IFAT の前身である FI は, 「知識の獲得や体系化」が設立時のスコープのひとつであった。フォーマルなデータであれば, 知識の獲得や体系化は比較的实现しやすい。例えば, ある人物に関する生年月日や著書などの客観的事実を様々な情報源から獲得し体系化する, といったことであれば Google がナレッジグラフとしてある程度実現している。しかし, 個々の事物に対する個人的な感情, 嗜好, 解釈といった表記方法が曖昧で時間や状況によって変化しうるものをどう扱うか, これらを例えばナレッジグラフのような知識体系と組み合わせてどう情報アクセスに使うのかなど, 検討すべき課題は多い。前節でも述べたとおり, IFAT で対象とされるデータはフォーマルなものから SNS などのカジュアルなものにシフトしているが, 従来のフォーマルなデータを扱う研究も, 新規のカジュアルなデータを扱う研究も, 情報アクセス技術の側面だけでなく, 情報基礎の側面にも焦点を当てた研究が少しでも増えてほしいと IFAT 研究会としては考えている。

参考文献

- 1) 福田 悟志, 難波 英嗣, 竹澤 寿幸. (2013) “論文と特許からの技術動向情報の抽出と可視化” 『情報処理学会論文誌データベース』, Vol. 6, No. 2, 16-29.

(難波英嗣・藤井敦)