



情報学基礎研究会

吉岡真治 北海道大学

情報学基礎研究会の概要

情報学基礎研究会は、情報自体が持つ基本的課題の明確化と関係する基礎理論の体系化を目標に 1986 年に設立され、以来、「情報自体を対象とする学問の基礎的研究領域」を活動範囲として、研究の振興と研究者の交流を目指して研究会活動を行っています。近年は特に、情報の有効活用という観点からの研究を行っています。

この背景には、大容量のデータストレージやインターネット環境の急速な普及に伴って、情報を取り巻く環境が大きく変わったということがあります。研究会設立当時の 20 年前において、大規模な情報は図書館などのごく限られたところにしか存在しませんでした。これに対し、近年では、インターネットを利用して、大規模な情報が非常に安価かつ容易に流通するようになりました。さらに、Web2.0 などのキーワードに代表されるように、以前ではきわめて限定的な情報発信能力しかなかった個人からの情報までも容易にアクセス可能になっています。

このような情報環境の変化に伴って、情報の有効活用という研究が目指す方向性が、限定的なところに存在する情報の有効活用という方向から、膨大な情報の中から有効な情報を探し出し、ユーザにとって理解しやすいかたちで提供するという方向へと変化してきています。

このような流れの中で本研究会では、基調テーマである、情報の分析・記述・表現やコンテンツの構築・利用・管理に関する基礎的な側面を基軸におきながら、技術的あるいはシステムの指向を持つ他研究会と補完することにより、情報環境の変化に対応した研究の展開を目指しています。

そのため、研究会を単独で開催するのではなく、他の研究会（データベースシステム研究会・自然言語処理研究会・デジタルドキュメント研究会・デジタル図書館ワ

ークショップ）との共催で開催することにより、幅広い視点からの議論が行える環境を提供しています。

また、コンピュータシステム領域のデータベースシステム研究会、電子情報通信学会データ工学研究専門委員会と共同で「情報処理学会論文誌：データベース」を発刊しています。本論文誌は、厳密なスケジュール管理に基づく迅速な査読を行うことにより、研究成果の速やかな公表の促進を目指しています。この結果、毎号 20 件近い論文投稿がある認知度の高い論文誌となっており、

NTCIR との連携

本研究会が扱う研究領域にはさまざまな研究がありますが、本稿では、本研究会との関係が深い情報アクセス技術に関する研究を行っている NTCIR (NII Test Collection for Information Retrieval)^{☆1}を中心に、近年の研究動向について簡単に紹介したいと思います。

NTCIR は、情報検索・テキスト要約・情報抽出などのテキスト処理技術を中心とする情報アクセス技術のさらなる発展をはかる評価ワークショップです。この評価ワークショップの目的は、個別の研究者の研究を比較可能なかたちで議論するための枠組みを構築する点にあります。

このために、NTCIR では、以下の 3 点の目的を設定して研究を行っています。

- 異なるシステムを一定の基準で比較するための基盤となるテストコレクション（課題群とそれに対する正解のセットを用意することにより、繰り返し検証可能な評価の基盤を提供する）の構築
- システム間の比較やアイデアの交換をできる研究者フォーラムの構築
- 情報アクセス技術の評価手法、評価指標に関する研究

本研究会におきましても、NTCIR におけるこれらの取り組みが研究会の目標と非常に近いことから、NTCIR 特集の研究会の開催、全国大会におけるパネルディスカッションなどを通じて、この NTCIR との連携をはかっています。

NTCIR の歩み

評価ワークショップは、米国の DARPA のプロジェクトとして開催された情報検索の TREC (Text REtrieval Conference)、情報抽出の MUC (Message Understanding

☆1 <http://research.nii.ac.jp/ntcir/index-ja.html>



◆ NTCIR-6 の様子

Conference) などに端を発し、さまざまなかたちで開催されています。

このうち、異なる研究成果の比較を可能とするテストコレクションの役割は、情報検索の分野においては、非常に高くなっています。特に、論文の審査などにおいて、大規模なテストコレクションによる評価をしていない研究が、採録されにくいといった状況にまできています。

このような、研究スタイルの変化に対応するように、日本でもテストコレクションの構築プロジェクト BMIR が本会のデータベースシステム研究会のワーキンググループで進められてきました。

NTCIR は、この流れを受けて、当研究会に所属する多くの研究者の参加のもとに、1997 年にプロジェクトとして発足しました。NTCIR ではワークショップを約 1 年半の間隔で開催しており、今年の 5 月に行われたワークショップで 6 回目を迎えることになりました。

NTCIR は、単言語・言語横断の情報検索の研究を中心に始まり、文書要約、質問応答といったテキスト処理技術の研究を含むかたちで発展しています。また、情報検索の研究では、Web 文書・特許文書などの特定のスタイルを持った文書に対する検索タスク(研究課題)などが実施されてきました。

参考までに、今年の 5 月に行われた NTCIR-6 ワークショップにおけるタスクを以下に示します。

- 言語横断検索(CLIR)
中国語・韓国語・日本語についての言語横断検索
- 言語横断質問応答(CLQA)
中国語・英語・日本語についての質問で、異なる言語の文書集合から質問への回答を生成
- 特許検索(PATENT)
特許文書を対象にした検索
- 質問応答(QAC)

日本語の文書集合から、理由を問うような質問への回答を生成

- パイロット課題・ワークショップ
意見分析・動向情報の要約と可視化

これらのタスクに対して、総計で 12 カ国から 85 グループが参加し、積極的なディスカッションが行われました。

また、近年のさまざまな利用者による情報要求の多様化に対応して、これらの情報アクセス技術の評価する手法についての関心も高まっています。このため、NTCIR-6 では併設ワークショップとして、このような情報アクセス技術の評価するための手法に関するワークショップが開催されました。

これらの議論を受けて、次回の NTCIR-7 からは、これまでのように個々のタスクを並行に行う形式ではなく、たとえば、質問応答に役立つ情報検索システムの評価というように、一連の情報アクセス技術の一貫として評価する方法が検討されています。

ただし、個別のタスクに関する参加はもちろん可能で、他の研究グループが作成したモジュールの組合せによる分析を可能にすることが考えられています。

むすび

本研究会は、情報の有効活用という考え方のもとに、情報環境を取り巻く環境の変化に伴って発展してきました。特に、情報爆発時代の大規模情報を扱うシステムの有用性の評価は、もはや、小さな研究グループによる個別の評価が困難な状況になってきており、研究者コミュニティによる評価基盤の確立が求められています。

本研究会では NTCIR などと連携することにより、これらの研究グループの相互交流の促進のための議論の場を提供したいと考えております。具体的には、来年の 1 月に、NTCIR における研究成果を中心とした研究会の開催を企画しております。詳細については、決まり次第、本研究会の Web ページ^{☆2}で広報いたしますので、この原稿を読んで NTCIR をはじめとする情報学基礎研究会に関連する研究領域に興味を持った皆様の参加をお待ちしております。

(平成 19 年 6 月 20 日受付)

☆2 <http://www.ipsj.or.jp/katsudou/sig/sighp/fi/>

吉岡真治(正会員)

yoshioka@ist.hokudai.ac.jp

1996 年東大・院・工博士課程修了。博士(工学)。学情センター、国立情報学研究所を経て 2001 年北大・院・工・助教授。2004 年同大・情報科学研究科助教授。2007 年より准教授。専門は情報検索への知識処理技術の応用。