1

筑波大学、新たなる挑戦

大学にとって、活力の源泉が研究の営為にあることは論をまたない。
一口に研究と言っても、すぐには日の目を見ない地道な基礎研究もあれば、時代のニーズに合致して脚光を浴びる花形研究もある。
筑波大学を代表するすぐれた研究にはどういうものがあるのか。そういう研究をされている方を執筆者として推薦してほしい、と各研究科長にお願いした。その答えがここにある。いずれもすばらしい研究だが、我が筑波大学のケイパビリティはむろんこれに尽きるものではない。

情報アクセス・システムおよびインターフェース機能の研究

石川徹也
図書館情報学系教授

はじめに

グーテンベルグの活字印刷技術の発明（1450年頃）により“本の出版”が始まった。その後約436年経って、1886年に著作権を国際的に保護する目的のベルヌ条約（Berne Convention：正式名称「文学的および美術的著作物の保護に関する条約」）が10カ国によって締結された（日本は1899年に加入）。以来、“商業出版”が始まり“出版産業”が進展する。ベルン条約はその後、録音・映画・放送等を含む多様な情報伝達メディアの著作権保護を目的に改正され、1971年のパリ改正条約等を経て、現在は、工業所有権等を含め、国連機関としての世界知的所有機関（WIPO）において著作権保護の条約策定・運用が行われている。

一方、科学技術の進展に伴い、情報伝達メディアは紙媒体（例：本とか雑誌）から機械可読媒体（例：フィルム）へ、そして現在は電子媒体（例：FDとかCD）へ、さらに通信媒体（例：データベースとかInternet）へと発展し、情報流通（情報の受発信）は“いつでも、どこにおいても”できるようになった。

個人および企業等組織の目的を効率良く達成するには、確度の高い情報の入手・活用が必要になる。当目的のために、多様にして大量の情報伝達メディアを、個人および組織において、日常的に維持・管理することは非常に困難である。そこで生まれたのが、図書館であり、また情報サービス機関（例：財団法人日本特許情報機構〔JAPIO〕）であり、情報サービス（例：天気予報とか交通情報）である。

図書館の成り立ちは「政（まつりごと）の記録（書）」を整理・保管する目的のために、古代アレキサンドリアの時代に創設されたと言われている。以来、商業出版物の流通により、広く出版物を利用に供するために、その提供活動は発展し、その活動を

‘学”として研究・教育する“図書館学”が生まれる。

また一方、人間の知的活動の成果を情報として効率的に流通・理解する目的のために、自然言語を解析・処理する技術の研究が計算機技術の発展と共に進展してきた。その成果はもはや、我々の生活に欠かせないワードプロセッサ（日本語においては仮名漢字変換システム）や、機械翻訳システム、さらには情報検索システム等を生み出している。これらの研究領域を“計算機言語学”(Computational linguistics) と呼称している。

筆者の研究領域は、自然言語処理に基づく多様なシステム機能の研究にある。

研究概要

現在では出版物等情報伝達メディアに関する情報（書誌データ）は、データベース化され、広く検索利用に供されている（例：公共図書館の蔵書データを携帯電話からも検索でき、利用予約できるようになってきている）。また、Internetの発展・普及により、あらゆる情報がフルテキストとして流通している。もはや、我々はInternet無くして、仕事を含め日常の生活はできなくなってきている。

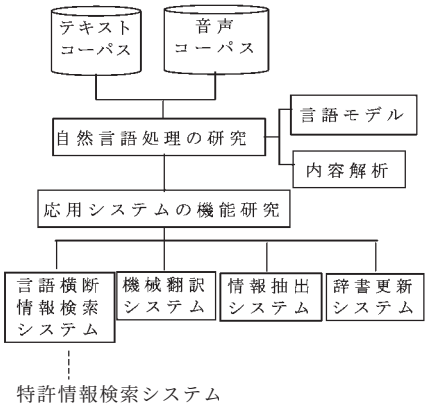
この情報入手・活用を支援するために、高精度の、しかも実用的な「情報アクセス・システムおよびインターフェース機能」の実現を目指し、システム化に必要な機能の研究を進めている。これらシステム機能は、基本的には人間系で行われてきた図書館サービス活動のシステム化といってよい（表参照）。

これら研究を現時点において、藤井敦助

図書館サービス機能（例） （図書館学における研究課題）	情報処理分野（特にInternet Computing）における研究課題
情報検索（Information Retrieval）	情報抽出（Information Extracting）　--特定情報の抽出・検索機能 部分検索（Passage Retrieval）　--ある必要部分の検索 言語横断検索（Cross-lingual IR）
索引語付与（Indexing）	自動索引（Automatic indexing）　--索引語の重要度判別
抄録作成（Abstracting）	自動要約（Summarization）
分類（Classification）	重要度分類（Clustering）　--検索結果の重要度判別・配列 カテゴリ化（Categorization）　--情報の事前体系への自動分類 フィルタリング（Filtering）　--情報の自動ふるい分け
選択的情報サービス （Selective Dissemination of Information Service）	情報推薦システム（Recommender System） --他者が有効とする情報を推薦する機能
レファレンスサービス （Reference service）	質問応答システム（Question-Answering system）

教授と大学院生、受託研究生等の協力の下に共同して取り組んでいる。

研究は基礎研究と応用研究に分かれる。基礎研究として、主にテキストデータおよび音声データを対象に言語解析モデルおよび内容解析の研究を行っている。応用研究として、解析結果を利用し、以下に事例として示す様に、多様なシステムの機能研究を行っている（図参照）。尚、音声データに関しては、藤井敦助教授が中心に進めている。



応用システムの研究（例）

1）多言語横断情報検索システムの研究

Internetの普及および種々の活動のグローバル化に伴い、情報流通は多言語に及んでいる。しかし、例えば、現時点のInternetを対象とする情報検索システムは、日本語キーワードを検索語として指示すると、日本語と英語のWebsiteを同時に検索するが、韓国語・中国語で記述されているWebsiteは検索しないといった問題点がある。同じ事項のことを、例えば韓国（語）では、中国（語）では、どの様に説明しているのだろうか、と思っても、韓国語、あるいは中国語で検索指示できないし、ましてや検索結果を判読できないことから、多くの人が諦めている（はずである）。

当研究は現時点において、日英中韓による検索語を相互に機械的に翻訳し、4言語によるWebsiteを同時に検索し、検索結果をユーザ言語に機械的に翻訳し、出力するシステム機能の研究を目指している。

当研究の成果は、現在、特許情報サービスを行っている企業からの受託研究を受け、産学連携研究の下に「多言語横断特許情報検索システムの開発」に応用し、既に一部製品化され好評を得ている。特にヨーロッパ特許情報機構（EPO）においては、日本語特許を英語で検索でき、特許公報等を英語で確認できることから、有効に利用されていると聞いている。

2）事典的情報抽出システムの研究

新聞を読んでいる時、テレビを見ている時、知らない“専門用語”に頻繁に出会う。その時、手元の辞書を、あるいは百科事典（辞典）を、さらには現代用語辞典等を調べたりしても、多くは発見できない。これら辞典等はスタンドアロンな情報伝達メディ

アであることから、タイムリーな更新はなされていないことによる。このことは冊子（出版物）の宿命である。

当研究はInternetに介して流通している多様な情報の中から事典的情報を抽出し、Webを巨大な百科事典のように使うことを目指している。

現在、当研究の成果を、情報処理技術者試験問題と対応する解説書を基に、試験問題に対する正解候補を抽出し、提示する応用システムの開発を行っている。このことから、当研究成果は、例えば学習者の回答に対して模範解答を提示することで、自己学習支援に応用できるものと考えている。

3）音声対応情報検索システムの研究

マルチメディア時代と言われて久しいが、文字情報の検索システムを除き、音声・画像情報の検索システムは未だの感がある。最近、音声認識研究の成果により、音声データを対象に研究できる環境が整いつつあり、音声データを対象とする多様なシステム機能の研究が盛んに行われだしている。

当研究において、例えば、講演後に、ある部分の説明を再度聞き確認（復習）するために、講演時に配布された文書（例えば予稿集）を検索し、関連する説明を、繰り返し聞ける音声対話検索（オンデマンド）システムを目指している。また、音声入力インターフェースを用いた柔軟な検索システムの研究も行っている。当研究は、独立行政法人産業総合技術研究所との共同で推進している。

当研究の成果は、例えば授業の復習（自己再学習）、コールセンター応答システム等に利用可能と考えている。

4）分類体系検索・情報分類システムの研究

図書館等において、情報伝達メディアの配架管理に、現在、国内においては「日本十進分類法（NDC）」が利用されている。国際的には国際十進分類表（UDC: Universal Decimal Classification）を利用している図書館が多い。十進分類法は全学問分野を10分野ずつ分け、体系化した、いわゆる知識の体系である。近年、UDCを言語に依存しない索引語・検索語ツール（Common Indexing / Retrieval Language）として、世界的に情報の分類・検索に利用する方式の研究がなされている（例：国のReal World Computingのプロジェクトにおける言語資源の構築において、ontologyデータとして使用された）。

筆者は、現在、UDC国際管理委員会（UDCC）委員に指名されていることもあり、UDCのデータベース化に伴い、当分類体系を、特に情報資源の管理に活用する研究を推進している。

4. その他システムの研究

研究の対象材料（言語資源）は主に論文

とか、新聞記事とかである。約 10 年前までは著作権の問題もあり、これら研究材料は電子化・頒布できず、データ（記述言語）の解析を机上でコツコツ行い、言語現象の規則を発見し、その範囲でシステム化し、評価用データも人手で作成し、評価実験を行ってきた。その後、例えば新聞社の協力の下に新聞記事を研究用に利用可能な形（大規模コーパス）として構築・頒布され、利用できる様になり、言語現象の解析は現在では、統計的な手法が中心になっている。この結果、当分野の研究が一段と活発になるのと同時に多様な研究がなされている。

この結果、現在未だ完成は見ていないが、大学院生の研究課題として、例えば自動要約（抄録）システム、情報のクラスタリング/カテゴライゼイション、質問応答、情報自動推薦システム等の研究を進めている。

おわりに

筆者の知識不足に起因することではあるが、日々、学生達から教えられること多である。最近では、韓国語のハングル文字表記においては、同音異義語（例：貴社、記者、汽車等）の表記は同じであることを教えられ、その自動識別を“ さてどうしたらよいか ”という大変な研究課題に着手することになった。この様に研究課題は次から次に現れる。

藤井敦助教授をはじめ、学生諸君、産学連携等の共同研究者に、種々感謝申し上げる次第です。

（いしかわ　てつや）