会議レポート

The Atmosphere of AIRS：情報検索の新しい国際会議の紹介

AIRS（エアーズ）とは？

「彼女，お高くとまっていやがる（She gives herself airs）」のように，air が複数形になるとろくなことがないようだが，本稿では AIRS（Asia Information Retrieval Symposium）という情報検索・情報アクセスに関する新しい国際会議について紹介する．その第 1 回は 2004 年 10 月に北京にて開催された [1]．

狭義の情報検索は文書・テキスト検索を意味するが，広義の情報検索は画像・楽曲検索や音声入力検索のように検索対象あるいは検索要求としてテキスト以外のメディアを扱う研究も含む．これに加え，テキストの自動要約，ユーザの質問に対する回答をずばり出力する質問応答など，情報洪水の中から有用な情報をすくい上げる技術を総称して情報アクセスと呼ぶことがある．AIRS はこれらのトピックを歓迎している．会議名には「情報とは空気と同様に我々の生活環境を満たす必要不可欠かつ多様なもの」という意味が込められているそうである．

情報検索の国際会議で最高峰といえば ACM SIGIR であろう [2]．毎年世界中から投稿があり，競争率は非常に高く，たとえば SIGIR 2004 のフルペーパー採択率は 58/267=22％であった．一方，SIGIR 協賛で開催された AIRS 2004 の採択率は 28/106=26％ で，主にアジアの研究者向けであるとはいえ，こちらもかなりの難関である．

「新しい国際会議」といっても，AIRS は突然ゼロから生まれたわけではない（AIRS did not appear out of thin air）．1996 年開催の IROL（Information Retrieval with Oriental Languages）および 1997 ～ 2003 年（2001，2002 年を除く）開催の IRAL（Information Retrieval with Asian Languages）という国際ワークショップの拡大版という位置づけである．IRAL がアジア言語を対象としていたのに対し，AIRS は対象言語も対象メディアも限定していない．たとえば自然言語処理の国際会議 ACL 2003[3] の併設ワークショップとして札幌で開催された IRAL 2003 のフルペーパー採択率は 13/25=52％ であったが，AIRS 2004 の採択率はこのちょうど半分であり，アジアにおける情報アクセス研究の競争率が急速に高まりつつあることが感じられる．この一因としては，国立情報学研究所が 1999 年より開催しているコンテスト型ワークショップ NTCIR（エンティサイル）により，アジア言語情報アクセス技術の評価環境が整ってきたことが挙げられる [4]．

AIRS 2004 の雰囲気

筆者は AIRS 2004 にプログラム委員および発表者として参加した．また，中華テーブルを囲んだランチタイム運営委員会にも飛び入り参加した（AIRS 2005 から運営委員を担当予定）．以下，簡単に AIRS 2004 について報告する．

AIRS 2004 は 10 月 18 日～20 日に北京友誼賓館（Beijing Friendship Hotel）という北京中心部からちょっと離れた巨大なホテルで開催された．参加者は 100 人強．1 日目にはホスト役を務めたマイクロソフト研究所（アジア）の研究者らによるチュートリアルおよびデモが行われ，2 日～ 3 日目には 8 つのフルペーパー・セッションが 2 つずつ並行で開催された．したがって各セッションの参加者は 50 名程度で，ちょうど情報処理学会の研究会のように，あまり気張らずに質疑応答が行える雰囲気であった（なお，AIRS 2005 では，参加者がより多くのセッションを聴講できるようにパラレル・セッションを減らす予定である）．また，このほかに 2 つのポスター・セッションが開催された．

以下，各フルペーパー・セッションで発表されたトピックをざっとまとめてみた．

- テキスト分類・クラスタリング，事象追跡（event tracking）
- テキスト要約（文抽出，文圧縮）
- アラインメント（テキスト対応づけ），言い換え，翻字（transliteration）
- Web 検索（リンク構造の利用，主要ページの検索）
- モバイル端末向け技術（特定地域向け検索，Web ページ要素の分類，画像検索ナビゲーションに関する招待論文 3 件）
- 適合文検出，地理的固有表現抽出，読解問題回答システム
- 検索モデル（言語モデルにおけるクエリモデルの推定など）

- その他（検索有効性評価尺度，文字認識による電子図書館コンテンツ作成など）

上記のようにトピックはバラエティに富んでおり，採択率を反映して個々の研究のクオリティも高いものが多かったように感じる．ただ個人的には，言語横断検索を謳ったフルペーパーが 1 件もないのは意外であった．また，モバイル関連の招待論文を除けば，まだまだテキスト処理ばかりという感も否めない．今後マルチメディア方面への展開が期待される．

フルペーパーを第一著者の所属に基づき地域別に分類してみると，韓国 7 件，中国本土 7 件，日本 6 件，香港 2 件，台湾 2 件，シンガポール 2 件，米国 1 件，タイ 1 件となる．ただし日本の 6 件のうち 4 件は在日外国人による論文であり，ちょっと寂しい．運営委員会でも，過去の IRAL に比べ日本からの発表者・参加者が少なかったことを心配する声が上がっていた．論文締切が第 4 回 NTCIR ワークショップ（2004 年 6 月 2 日〜 4 日，国立情報学研究所）の直後であったことが影響したのかもしれない．あるいは，AIRS という名前が耳慣れなかったのかもしれないと思い，今回，宣伝のために筆をとった次第である．

AIRS 2004 の会議録には，3 件の招待論文を含む 28 件のフルペーパーと 38 件のポスター論文が収録されている．また，フルペーパーは Springer-Verlag から Lecture Notes in Computer Science シリーズとして別途出版される．研究者にとって魅力的なこの出版制度は来年以降も継続予定である．

AIRS の今後

次回の AIRS は 2005 年 10 月に韓国済州島で開催される [5]．魅力的なロケーションなので，情報検索屋ならば思わず論文が書きたくなってしまうであろう．自然言語処理の国際会議 IJCNLP[6]（こちらも 2004 年発足）と連続開催の予定で，論文投稿締切は 4 月 15 日頃．マルチメディア検索のセッション，ビデオ・マイニングの特別セッション，および展示セッションを新たに企画する予定である．ぜひ日本からの投稿および積極的な参加をお願いしたい．

最後に，気が早いが AIRS 2006 はシンガポールで開催予定である．

謝辞　本稿執筆にあたり AIRS 運営委員会から多くの情報およびご意見をいただいた．ここに感謝する．

参考 URL
1）AIRS 2004: http://research.microsoft.com/nlc/airs2004/
2）ACM SIGIR: http://www.acm.org/sigir/
3）ACL 2003: http://www.ec-inc.co.jp/ACL2003/
4）NTCIR: http://research.nii.ac.jp/ntcir/index-ja.html
5）AIRS 2005: http://www.airs2005.org/
6）IJCNLP 2005: http://www.afnlp.org/IJCNLP05/

（酒井哲也 /（株）東芝 研究開発センター）