

J-STAGE に登録されている本学会論文誌のアクセス統計の解析(4)

中野 英彦

兵庫県立大学名誉教授 (671-2280 姫路市書写 2167)

【緒言】

本学会の論文誌である The Journal of Computer Chemistry, Japan 誌 (以後本誌という) は、創刊号より通常の紙媒体で出版するとともに、オンライン公開においては学会のホームページ上と、JST (独立行政法人科学技術振興機構) によって運用されている J-STAGE (科学技術情報発信・流通総合システム) 上に全文を並行して公開していた。その際、全文の提供形態としては、学会ホームページ上では HTML 形式および PDF 形式を、また J-STAGE 上では PDF 形式のみを使用していた。しかし、PDF 形式での提供を重複して行うことは、サーバーなどの資源ならびに作業の面において無駄を生じるだけでなく、J-STAGE から提供されるアクセス統計に利用者からのアクセスの情報が一部しか反映されないというデメリットがある。また、本年から J-STAGE での公開において、出版に先立って「早期公開」を行うことにしたので、これまでオンライン公開用の PDF ファイルが完成した時点で印刷体出版に先立って公開するという、学会サーバー上での公開のメリットがなくなった。そこで、本年から PDF による公開を J-STAGE 上でのみ行うこととした。しかし、学会サーバー上でのオンライン公開を完全に廃止する事はせず、論文誌の目次、英文要旨、和文要旨、および全文の HTML ファイルによる提供を継続し、全文の PDF を読みたい利用者に対しては J-STAGE 上のファイルへのリンクを提供する事により対応した。

演者は過去 3 年間の秋季年会⁽¹⁾において、J-STAGE から提供される統計データに基づく解析結果を発表したが、今回は特に上述した本年からの変更によるアクセス傾向の変化についての情報を提供し、論文投稿者の参考に供したい。

【方法】

J-STAGE から提供されるデータは「アクセス統計レポート」と「加工済みログ」の二種類である。前者は、J-STAGE のシステムで取得したサーバーのログファイルを、アクセス統計プログラムで集計した結果であり、「記事別レポート」、「国別レポート」、「ドメイン別レポート」などとして整理されており、表計算ソフトを利用して解析可能であり、前回までの解析においてはこれらのみを利用したが、今回はさらに「加工済みログ」についても利用することとした。そのためには、J-STAGE から提供されるデータを独自に処理するため、簡単な処理プログラムを Visual Basic によって作成した。

【結果】

1. 全文 PDF ファイルのダウンロード数

記事別レポートを見れば、それぞれの記事 (論文) が、書誌事項 (要旨など) 全文 PDF ファイル、引用文献情報などのファイルごとに、1 ヶ月に何件のアクセスがカウントされたかがわかる。それらの内で、論文の内容を全文取得するという意味で、最も重要と思われる全文 PDF ファイルのダウンロード数を中心に解析を行った。

その際、前回までの解析においては、「アクセス統計レポート」の中の「記事別レポート」のみをデータとして利用していたが、今回「加工済みログ」を精査したところ、統計データ取得月によっては特定日時の短時間内に、特定の IP アドレスからほとんど全論文についてのアクセスが集中して行われている例が含まれている場合があることが判明した。これは、特定のサイトが本誌の全データを取得するために組織的に行っているものと判断した。J-STAGE によって統計データを取得する際にも、このような例は排除する事になっているはずであるが、その際に個別的に精査して判断するのではなく、何らかのアルゴリズムによって機械的に行っているものと推定され、場合によってはその判断に漏れているものが混入しているものと思われる。そこで、今回からは、

「加工済みログ」を精査した結果、特定のサイトからの意図的な大量データ取得と判断されるデータについては、「記事別レポート」と突合せて、記事別レポートからも排除する事とした。

図1に、全論文(第1巻第1号より第7巻第2号までの143論文)についての2008年1月より7月までのダウンロード数の月平均の分布をアクセス数の多い順にソートしたものを示す。

最もアクセス数の多い論文は30回/月であり、少ない論文は0.4回/月であり、平均値は4.1回/月である。これは昨年(17-0.8,平均3.4/月)⁽¹⁾比べて増加しており、特にアクセス数の多い論文においてその増加率が顕著である。

図において、淡色で表示されている8論文は、本年になって新方式(PDFファイルをJ-STAGEのみに搭載する)を適用して以後の論文であり、高アクセス数の論文の大半を占めている事がわかる。この結果は、これまで本学会のサーバーからダウンロードされていた数に加わる事になったためと思われる。

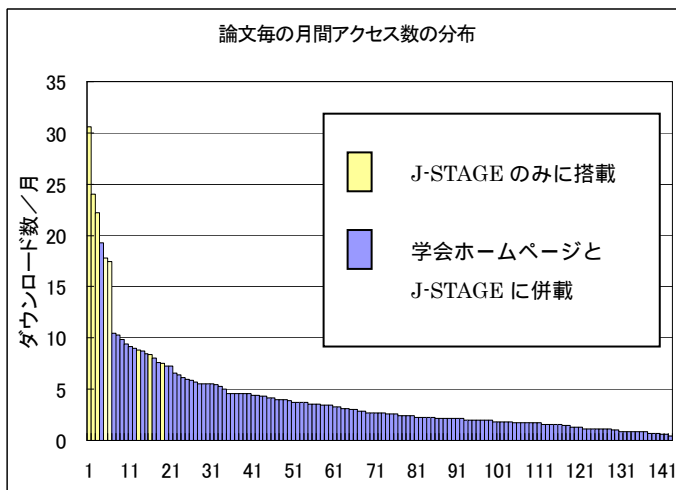


図1 論文毎の月間アクセス数の分布 (2008年1月-7月)

2. 国別のアクセス割合

前回までの解析では「アクセス統計レポート」の中の「国別レポート」を利用していたので、論文記述言語によるデータが含まれておらず、和文論文と英文論文における国別のアクセス割合を判別する事ができなかった。今回、「加工済みログ」のデータを併用する事により、こちらで把握している論文記述言語と付き合わせる事により、英文論文と和文論文の違いを示す事が可能となった。全論文に対しては、ほぼ2/3が国内からのアクセスであるが、英文論文に対しては国内からが1/4であり、海外からのアクセスが多数を占めている。

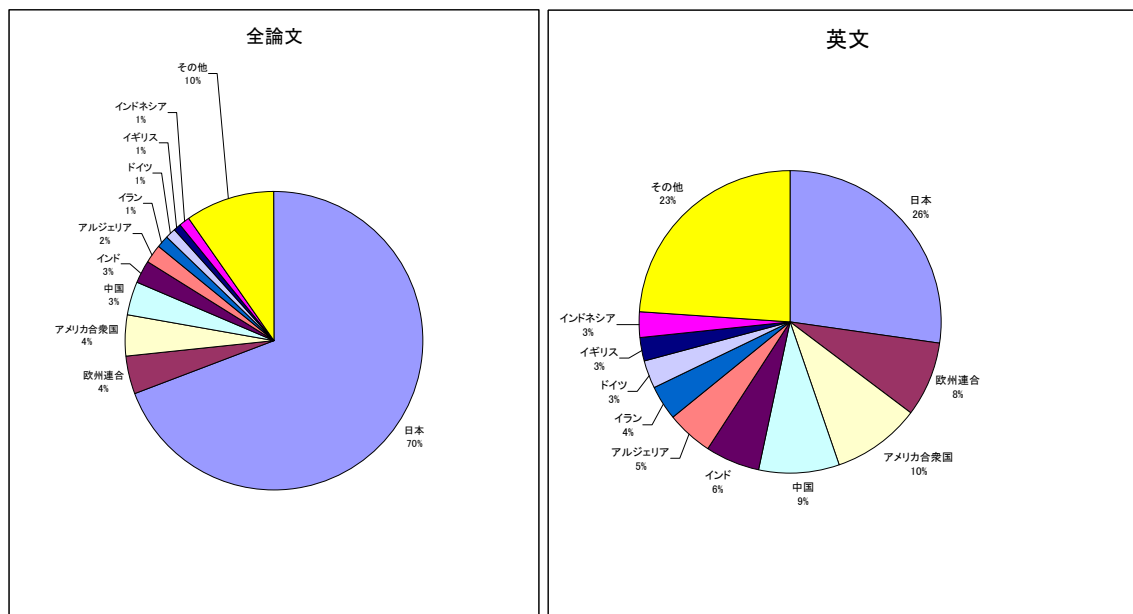


図2 全論文および英文論文に対するアクセスの国別割合 (2007年7月のデータ)

参考文献

(1) 中野英彦、日本コンピュータ化学会 2007 秋季年会講演要旨集、2P14 (2007)