1 新しい応用の動向

～総論～

正会員 村岡 洋一[†]

キーワード

インターネット, グリッド, Web情報アーカイブ, セマンティックWeb, テキストマイニング

1. ま え が き

インターネットの誕生は、今から30年以上の昔である1969年のARPANET誕生にさかのぼる。そして、この30年間、電子メール、ネットニュース、WWW (World Wide Web)、P2P (Peer to Peer)、グリッド (Grid) といった革新的なアプリケーションの誕生に伴って、インターネット利用者数は増大を続けている。また、2002年末には、我が国における家庭へのインターネット普及率が全世帯の8割を超える¹⁾など、研究者のためだけのインターネットから一般ユーザのためのインターネットへと変貌してきている。

本特集では、このように新しい応用の開拓が急速に進みつつあるインターネットについて、「計算」と「情報」という二つのキーワードを取り上げ、最新の技術動向を紹介する。

まず、「2. 「計算」応用の開拓」では、グリッド技術を取り上げる。グリッド技術とは、ネットワークで接続される多数のコンピュータの資源を活用することで、大規模な処理能力を実現するための技術であり、「Power Grid」(送電網)が語源となっている。今や、計算機は、所有する時代から、電気、水道、ガスのように必要な分だけ利用する時代へと変貌しようとしており、グリッド技術は、次世代の社会インフラを支える技術といっても過言ではない。

グリッド技術の応用としては、世界中に点在する何百万台ものパソコンの余剰能力を活用して、宇宙からの電波を分析し地球外生命体を発見しようとするSETI@Homeプロジェクト²⁾や、タンパク質の折り畳みを予測し、アルツハイマー病、狂牛病などといった多くの病気の解明に役立てようとするFolding@Home³⁾などがあり、専門家だけでなく多数の一般ユーザが参加している。これらは、技術的にはメタコンピューティングと呼ばれる。2章では、これらのメタコンピューティングを含め、グリッド技術の最新動向として、グリッドポータル、データグリッド、アクセスグリッドについて紹介する。

執筆は、グリッド研究に関する国内最大の団体である「グリッド協議会」⁴⁾の設立者であり、国内のグリッド研究

の中心である産業技術研究所でグリッドセンターのセンター長をされている関口智嗣氏のグループにお願いした。

続いて「3. 「情報」の応用の開拓」では、全世界のWeb情報アーカイブ構築プロジェクトを取り上げる。

インターネット上でWWWを通じて発信される情報量は、テキストのみで約80億ページと推定されている(2003年8月時点)。これは、全世界の人が各人平均1.3ページの情報を発信しているのと同値であり、膨大な情報量である。Webページ1ページあたりの容量が平均15KBであると仮定し単純計算すると、テキストだけでも120TBものデータがインターネット上で公開されていることになる。さらに、この量は毎年増え続けると共に、常に更新され続けている。そして、いわゆる検索エンジンなくしては、インターネット上の情報を有効活用することができなくなっている。

3章では、このような膨大な情報をどのようにして高速に収集し検索可能な状態にするかという点について、本分野で長年研究を続けている早稲田大学の山名早人氏に執筆をお願いした。

4章は、3章に引き続き「「情報」の探索」という観点から、膨大な情報をどのように探索するかについて、近年ホットな話題となっているセマンティックWebを取り上げる。

セマンティックWebは、W3C (World Wide Web Consortium)⁵⁾のディレクタであり、WWWの発明者として知られるTim Berners-Lee氏が提案する次世代Webのビジョンである。セマンティックWebは、Webの情報を検索・処理しWeb上でのさまざまな問題解決を可能にすることを目指している。セマンティックWebが実現されれば、サーチエンジンの適合率(検索精度)を格段に向上させることが可能になると共に、QA検索などの意味内容を踏まえた質問応答システムの実現が容易になる。QA検索とは、ある質問に対して該当するWebページを表示するのではなく、Webページに掲載されている情報を基に、直接回答を作成しユーザに表示する検索方式である。

執筆は、INTAPセマンティックWeb委員会でご活躍されている富士通研究所の津田宏氏にお願いした。

最後に5章では、「「計算」と「情報」の融合」という観点から、膨大なテキスト情報の中から有用な規則を発見するテキストマイニングを取り上げる。

[†]早稲田大学 理工学部 コンピュータ・ネットワーク工学科
"Trend of New Applications -Introduction-" by Yoichi Muraoka (Waseda Univ., Tokyo)

テキストマイニングは、データマイニングの対象をテキストデータに特化したものである。例えば、大量のビジネスデータ(テキスト)から、ビジネスに有用な規則やトレンドを自動抽出する技術がテキストマイニングである。テキストマイニングに関する技術は、各企業・研究所で研究開発が進められており、すでにさまざまな会社から製品が販売されている。しかし、その分析速度やスケーラビリティは充分ではない。特に、3章で取り上げたような膨大なWebデータ(テキスト)を対象とした場合、現在のシステムで対応するのは困難である。このような背景のもと、5章では、テキストマイニングの高速処理に焦点をあてている。

執筆は、テキストマイニングの高速処理をインデックス方式の高性能検索ソフトウェアを用いて実現したアクセラテクノロジー社の萩原純一氏と進藤哲也氏にお願いした。

次章以降では、これらの技術についての最新技術動向を紹介する。インターネットが誕生して30年間の間に、WWWの誕生、グリッドの誕生といったさまざまな新しい応用が考えられると共に、インターネットは、今や生活に欠かせないものとなっている。インターネットの未来を予想することは困難であるが、本特集で取り上げた技術に、近未来のインターネットの姿を見い出すことができるものと期待する。

最後に、ご多忙中にもかかわらず貴重な寄稿をいただき

た執筆者の皆様、本特集号の企画、立案にご尽力いただいた編集委員の皆様、編集作業等いろいろとお世話になった学会事務局の皆様にご心から感謝いたします。

(2003年10月28日受付)

〔文 献〕

- 1) 総務省：“平成14年度版通信使用動向調査”(Mar. 2003)
- 2) SETI@Home, <http://setiathome.ssl.berkeley.edu/>
- 3) Folding@Home, <http://folding.stanford.edu/>
- 4) グリッド協議会, <http://www.jpgrid.org/>
- 5) W3C, <http://www.w3.org/>

むらおか よういち
村岡 洋一 1965年、早稲田大学理工学部卒業。1971年、アメリカ イリノイ大学電子計算機学科博士課程修了(Ph.D.)。同大学Visiting Research Associate、日本電信電話公社(現NTT)電気通信研究所などを経て、1985年より、早稲田大学理工学部教授。2002年より、早稲田大学副総長、理工学部教授、COE「プロダクティブICTアカデミアプログラム」拠点リーダー、情報処理学会副会長、電子情報通信学会フェロー、情報処理学会フェロー、GRID FOR UM Advisory Committee Member、IEEE Computing誌Editor、International Journal on Parallel Programming誌Editorなども務める。電子通信技術審議会、産業構造審議会専門委員、学術振興会委員、科学技術会議専門委員など各種委員会委員を歴任。正会員。