

ウェブとデータベースによる情報配信

成 凱
Kai CHENG

九州産業大学 情報科学部 社会情報システム学科
Faculty of Information Science, Kyushu Sangyo University
chengk@is.kyusan-u.ac.jp, <http://www.is.kyusan-u.ac.jp/~chengk/>

1. はじめに

私の研究室では、主にデータベースとそれを駆使した情報配信システムについて研究開発を行っている。これまでの研究活動は以下の二つの方向に大別できる。

- ウェブとデータベース (WebDB): ウェブ上での情報配信を目的とするデータベース技術に関する研究。研究テーマ: ウェブキャッシュ、ウェブウェアハウス、利用履歴によるウェブ検索の精度向上、学習教育支援システム
- ネットワークとデータベース (NetDB): ネットワークを經由して流れてきたデータ、いわゆる「ストリームデータ」の解析とその利用に関する研究。研究テーマ: 半永続データの自己組織化管理、ストリームデータ管理のための効率的アルゴリズムの開発、センサーネットワーク及び IC タグの応用システムの試作

本稿では、それぞれの研究方向における具体的な研究内容について述べる。また、卒業研究などで実際に行っている研究テーマをいくつかピックアップし紹介する。

2. ウェブとデータベース

ウェブはインターネット上の情報配信サービスのひとつである。ウェブ上では情報をウェブページの形で配信している。ウェブページは HTML というマークアップ言語で記述されるため、HTML 文書とも呼ばれる。すべてのウェブページは URL という唯一のアドレスをもち、それを使って世界中のどここのページでも特定できる。画像や映像を含むマルチメディア情報を埋めることができるため、幅広い種類の情報を配信することができる。さらに、ページにリンクを埋めることによって、ページ間に関連付けができる。

ウェブを使うことによって情報配信のコストが大幅に低下するとともに、誰でも簡単に情報を発信できる利便性と、膨大な情報源として世界のどこからでも簡単にアクセスできる効率性を持つことで、印刷術やマスメディアに次いで新しい情報革命を起こしているといわれている。

2.1 ウェブキャッシュ/ウェブウェアハウス

ウェブの普及に伴い、ウェブの利用者数と一人ひとりの利用頻度がともに急増している。限られたネットワーク帯域幅は急増するトラフィックに耐えられなくなると、ネットワーク通信ができなくなる恐れがある。ネットワークのボトルネックを解消するために、キャッシュを用いて、同じページに対するリクエストはできる限り近くのキャッシュを利用するように工夫する必要がある。キャッシュの利用率を高めるために、利用パターンを予測しこれから利用しそうなコンテンツを特定しキャッシュに保存する研究がよくあった。

しかし、実際に利用状況を解析してみると、ウェブの利用者は同じコンテンツを何度もみるより、新しいコンテンツを探すことが多いので、利用パターンは簡単に予測することができない。ここで、我々はコンテンツの意味上の関連を考慮したキャッシュ手法を提案した [1]。例えば、夏だと、海水浴、花火大会といったコンテンツに辿っていく可能性がより高いため、それらのコンテンツを優先的にキャッシュしておく、キャッシュの利用率(ヒット率)を改善できる。

さらに、モーアの法則により、キャッシュに使えるハードディスクは価格が安くなるとともに、容量による制限も緩和しつつある。そのため、キャッシュに残すべきコンテンツを選ぶより、キャッシュコンテンツをデータベースのように利用することが重要となってきている。キャッシュの一般化として、ウェブウェアハウスを考案した [2]。ウェブウェアハウスで実現可能な機能として、高度のデータ管理機能、意味関連による利用パターン発見機能、コンテンツ推薦機能、協調サーチ機能を考案している。

2.2 利用履歴によるウェブ検索の精度向上

ウェブ上で膨大な情報が利用できるものの、検索して必ずしも自分の興味にあう情報が出てくるとは限らない。そこで、利用履歴を解析して利用者の多いページがわかれば、そこから出てきたページを検索結果の上位に示すことで、より利用者に興味に近い情報を得ることができる。例えば、情報科学部の教員のページから検索すると、同じキーワードを含むページが複数あった場合は、これまでの利用履歴からみて、人気の高いページに高いラン

クをつけて検索結果リストの上位に示す。このように、ほかの利用者の利用履歴を推薦として利用することで検索の精度を高められる。利用状況をグラフで表示することで、情報の可視化にも適用できそうなので、現在、その実現を検討している。

2.3 学習教育支援ウェブシステム

ウェブ技術は最初開発されてからずっと進化しつづけている。最初は静的なページを中心とする静的コンテンツだったが、サイトの規模が大きくなればなるほど、静的なコンテンツとして管理することが困難になる。データベースを利用して動的なコンテンツ生成に変わりつつある。さらに、ユーザとのやり取りを加えると、ウェブインターフェースを介してさまざまなサービスを提供するようになっている。特に、最近 Ajax、Rails といった技術や開発環境の普及に伴い、ウェブはさまざまなアプリケーションのインターフェースに変わる流れが加速している。

この流れのなかで、私の研究室でも、身近な事例を取り上げて、学習教育を支援するウェブシステムの研究開発も行っている。以下にその代表的なものを紹介する。

1. 出席登録システム

パソコンを使う授業を前提に開発したこのシステムでは、学生一人ひとり自分のアカウントでログインし、出席登録をはじめ、出欠状況確認、レポート提出、小テスト成績確認など、すべてウェブ上で行える。座席上に学生が選んだ画像を表示したり、学年暦や時間割を確認したりすることもできる。機能としてはかなりシンプルだが、実際に使ってもらった学生の反応は上々である。これからもっと学生の力を借りて、新しい機能を追加していくと考えている。

2. 授業体験 on the Web

高校生は大学の潜在顧客である。大学の授業の進め方や学習の内容を彼らに体験してもらうのは、大学への理解を深めるとともに、進路選択の参考に供するものとして重要である。従来の授業体験では、参加者に物理的な教室に集まってもらわないと授業を受けることができない。しかし、教室の収容人数や授業の時間割などによる制限が多く、効率的かつ効果的に実施することが困難である。

インターネットの普及や高校で情報が必修となっている背景に、情報技術の専門家を育てる学部として、専門分野の魅力を高校生に理解してもらうため、情報技術を活用した体験教材を開発することが望まれる。そこで、我々はウェブ技術を利用してブラウザがさえあればいつでもどこでも大学の授業を体験参加できる「授業体験 on the Web」(図 1)を開発することにした。

このシステムは九州産業大学情報科学部で 4 年間運用してきた講義記録システムに録画されている映像データを利用し、ウェブでの出席登録、座席レイアウト確認がで



図 1 高校生向けの授業体験 on the Web システム

きるほか、映像で講義を聴講することも可能である。参加者には遠隔地に居ても参加者同士が同じ教室で受講しているように、臨場感を与えることができる。オープンキャンパスや大学見学の場で高校生に実際に利用してもらうことで、このシステムの有用性を検証した。

3. 研究室わけ希望調査 Web システム

情報科学部では、学生は 3 年後期になると、各研究室に配属され、社会情報システム学・知能情報学演習を受ける必要がある。この演習のクラスは原則として卒業研究のクラスになるので、学生にとってかなり重要な選択のひとつである。学生の希望状況、成績、研究室の定員などを総合的に考慮して、希望調査を 2 つの段階にわけて実施することは学部で決められている。調査の実施から、調査結果のまとめ、配属結果の決定まで、情報の収集と管理が重要なので、データベースを使い、ウェブインターフェースを介して実施することにした。

このシステムを使って、調査に参加する学生は希望を出した時点で、各研究室の定員、希望者数などをその場で確認できる。配属決定の手間を省くことができる。このシステムの詳細については本誌 4 巻 1 号 [3] をご参考ください。

本研究室は社会情報システム学科に所属している。そのため、学科名にあるような社会情報システムはいったいどんなシステムなのか常に考えている。私は複数の人間からなる組織やコミュニティを効率よく機能させるための情報システムと理解している。組織が効率よく機能するために組織内部における情報共有と外部に対する情報配信が重要である。大学はひとつの小さな社会なので、上記のシステム開発は社会情報システムへの理解も深まっ

たと感じている。

3. ネットワークとデータベース

データベースはウェブと緊密に関連しているだけでなく、ネットワーク分野との関連も強まっている。特に、センサーネットワーク、RFIDなどで、人間が関与せずに自動的に情報を収集できるため、既存のデータベース技術だけでは対応できなくなる。このようにネットワークを経由して流れてきたデータはストリームデータと呼ばれ、その解析と利用はますます重要となってくる。本研究室では、これについて、以下のテーマで研究を行っている。

3.1 半永続データの自己組織化管理

ストリームデータと従来のデータベースのデータと比べると、大きな違いはデータの永続性にある。従来のデータベースでは、データは一度蓄積されたあと、明示的に削除や変更をしない限り、永続的に蓄積され必要な時いつでも調べられると想定されている。

一方、ストリームデータは高速で無限に到来し続けるため、すべてのデータを格納することは現実的には不可能であるし、格納されてもリアルタイム性がないため、あまり役に立たなくなってしまう。従って、ストリームデータは保存せず一度しか見ることできないデータである。

しかし、キャッシュデータのようなデータは一定の期間において格納されるが、期限が過ぎたり、ほかのデータに置き換えられたりすると、格納されなくなる。我々は、このようなデータを半永続データと呼び、その性質の解明とモデル化を図っている。半永続データの特徴は(1)データに寿命がある。寿命が終わるとデータが無効になるか削除される。(2)データそのものと、データに対するトラフィック(アクセス頻度、最近の人気度)と合わせて保存する。そのため、利用状況を考慮した自己組織化管理が可能になる。半永続データの研究によってデータ工学の新しい分野の形成が期待されるとともに、利用パターンを反映した情報配信システムとして、実用化も期待できる^{*1}。

3.2 センサーネットワークとRFIDの応用

近年、センサーネットワークやRFIDといった新しい情報の取得・転送技術が現れて、注目が集まっている。

RFIDは物流管理など流通業界でバーコードに代わる商品識別・管理技術として研究が進められている。またBSE問題や偽造表示による品質管理、トレーサビリティ(商品履歴管理)要求の高まりにより、RFIDにより一層の期待が集められている。RFIDとはICタグを使った無線通信による識別技術であり、数センチメートル程

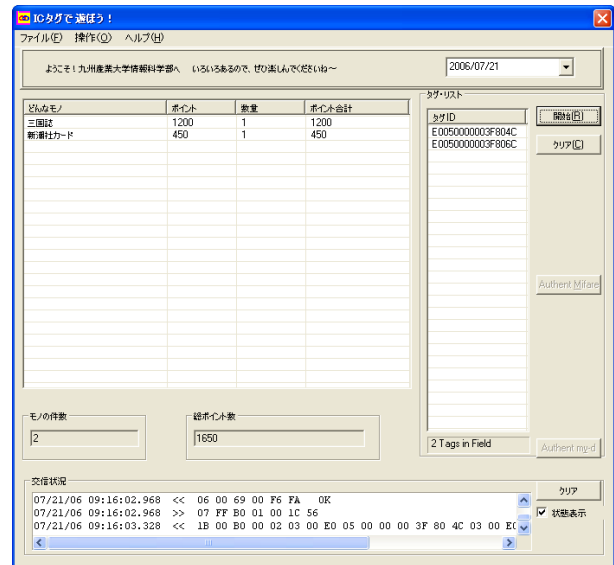


図2 ICタグで遊ぼう

度のタグ(RFIDタグ)にデータを記録し、電波や電磁波でリーダ(RFIDリーダ)と通信を行う。RFIDの特徴として、非接触でのデータの読み取り、書き換えが可能であり、アンチコリジョン機能を用いると一度に複数のICタグの読み取りを行うことができる。このようにRFIDを用いることで、ネットワークを介して個体識別する技術が実現されようとしている。RFIDを用いることにより、ネットワークを介して、オブジェクトの位置情報などの状態が取得可能となる。

1. ICタグで遊ぼう!

RFIDは限られたメモリしか持たずタグIDしか蓄積できないため、ネットワークを介してデータベースと接続する必要がある。RFIDの可能性とそれに伴う問題を理解するため、本研究室では、いくつかの実験システムを開発している。図2の画面は、本研究室と牛島研究室と共同で行ったRFIDの実験のひとつ「ICタグで遊ぼう!」を示している。この実験システムでは、タグ付けのカードからIDを読み取り、そのIDに関連する情報をデータベースを調べ、ディスプレイ上で表示する。カードの表に関連情報が前もって印刷してあるが、それを見なくても情報の内容がわかる。このシステムは今年度第1回目のオープンキャンパスで出展し来場者の興味を惹いた。

2. RFID無人レジシステム

本研究室で実現したもうひとつの実験システムはRFID無人レジシステムと呼ばれ、スーパーマーケットなどのレジシステムの無人化を目指している。RFIDを使うことによって商品に付いたタグからどの商品を識別し、関連データをデータベースから調べ、商品を買物ハゴにいったままで、清算ができる。

3. イベント運営支援

センサーネットワークとは、超小型の無線装置を内蔵したセンサーが自動的に周囲の状況を認識し、認識した

*1 文部科学省科学研究費補助金、萌芽研究(2):半永続データの自己組織化管理に関する研究 平成17~18年

情報をもとに適切な動作を行ったり、無線によりセンサー同士で情報共有を行ったりするネットワークのことである。センサーネットワークが実現すると、利用者やモノの状況を自動的に認識し、利用者の状況に応じたさまざまなサービスをリアルタイムに提供することも可能になる。

本研究室では、センサーネットワークの実験システムとして、MOTE というセンサーネットワーク開発キットを使って、イベント運営支援システムの研究開発を行っている。大型のイベントを運営するには会場の状況の把握と、運営スタッフの間の情報共有は重要である。また、イベント実施する前までの役割分担や、実施開始後の担当者変更などを順調に実現するため、データベースを用いた情報管理が重要である。開発中の実験システムでは、これらの機能の実現を行っている。

4. 卒業研究テーマ例

冒頭から述べたように、本研究室では、ウェブとデータベースによる情報配信システムについて、研究開発を行っている。そのため、データベースの知識とウェブの知識は重視している。卒業研究では、ソフトウェアの開発がメインとなっているが、パソコンの中身の確認から、OS のインストール、サーバソフトのインストール・設定、アプリケーションの実現まで、一人でできることが望まれている。ハードウェアの設定からソフトウェアの作成まで一通り経験することによって、実践力を高めるために重要だと考えている。

研究テーマは 4 年次に入ってから 2ヶ月後頃に個別相談をした上で決める。学生からやりたいことがあるなら、卒業研究テーマとして相応しいと判断すれば、研究テーマはそれに決定する。自分でどうしても決められない場合は、教員からいくつかの候補テーマを提供し、そこから選ばせる。一人ひとり違うテーマでやっているのだから、大変かもしれないが、研究は自分のペースで行うことができるメリットもある。今年は以下のようなテーマを決定している。

- (1) 情報科学部施設ナビゲーション Web システム
- (2) Ajax による「情報科学部授業体験 on the Web」の操作性の改良
- (3) 「研究室分け希望調査 Web システム」における研究室決定手続きのシミュレーション機能の実現
- (4) ソーシャルネットワークによる履修支援
- (5) RFID を用いた研究室在室状況管理
- (6) 利用状況解析によるウェブコンテンツのホットスポットの検出と可視化
- (7) SQLite による軽量的でポータブルな講義 HP 運用管理ツールの開発
- (8) 出欠状況データの統合とその利用手法の提案

学生から提案があり教員の意見を加えたテーマの一つは「ソーシャルネットワークによる履修支援」である。学



図 3 履修支援ソーシャルネットワーク

部で開設されている科目に対し、同じ学部で学んでいる学生同士、担当している教員、そして卒業生の間、意見や情報を交換するためのコミュニティソフトの開発を目指している。時間割、スケジュール、友達リンク、講義に対する口コミ情報、オンラインアンケート調査などの機能の実現を計画している。

5. ま と め

本研究室で行っている主な研究内容と、平成 18 年度卒業研究で取り組んでいる研究テーマについて紹介した。

◇ 参 考 文 献 ◇

- [1] K. Cheng and Y. Kambayashi. A Semantic Model for Hypertext Data Caching, In Proceedings of the 21st International Conference on Conceptual Modeling (ER 2002), Tampere Finland, 7-11 Oct. 2002. Springer. LNCS 2503 pp. 276-290
- [2] Y. Kambayashi and K. Cheng. Capacity Bound-free Web Warehouse, In Proceedings of the First Biennial Conference on Innovative Data Systems Research (CIDR 2003). Edited by M. Stonebraker, J. Gray and D. Dewitt. Asilomar, CA, USA, 5-8 Jan. 2003. pp.47-57.
- [3] 成凱, 仲隆, 宮崎明雄. 研究室分け希望調査支援 Web システム 九州産業大学情報科学会誌 4 巻 1 号 (2005 年 9 月) pp.34-38
- [5] 梅田 望夫. ウェブ進化論—本当の大変化はこれから始まる ちくま新書 5 8 2