

情報検索システムの土木分野での利用方法の研究

学生氏名 永井達也
指導教員 皆川 勝

1. はじめに

情報検索の歴史は、コンピュータの発明とほぼ同時に始まった。そのシステムの最も基本的な機能は、「データ群」に対して、人間が「検索要求」を出し、コンピュータが「照合」処理を行って、その結果として「該当文章」を得るというものである。近年のインターネットの普及により、情報検索システムは全く新しい形態へと変化し、今後建設省が進める建設 CALS/EC の普及及び管理に多大な影響を与えると考えられる。本研究は、情報検索システムがどのように土木工学の分野で有効に利用できるかを探ることを目的として、いくつかの検索システムを試作した。

2. 建設 CALS/EC

CALS (Continuous Acquisition and Life-cycle Support) は米国の軍の後方支援部門の膨大な紙の資料を電子化することによって、そのコストを大幅に削減し、その情報を共有化して再利用することにより、組織に大なる効果をあげた活動に始まっている²⁾。この技術をもとに、EC (Electronic Commerce : 電子商取引) と組み合わせた建設分野での利用を目的とした活動および概念が建設 CALS/EC である。これは、限られた公共事業の財源を有効に活用して公共資本整備を充実させるため、及び本格的な高齢化社会到来に備えるための、建設コストを一層縮減する有効な方策の 1 つとしてとらえられている。しかし、実際の建設 CALS/EC の実現には、設備機器・ソフト他大きな投資を必要とするばかりでなく、解決しなければならない課題も多い。

3. 全文検索システム Namazu

インターネットの加速度的な進展に伴い、各サイトの情報の量が増加し、求める情報の検索、取得が困難になるといった状況も発生してきた。このような事態にいち早く対応した一部の企業サイトでは、トップページに検索サービスをつけてこれを回避しようとする動きが見られるようになった。ここで用いられているのが、全文検索ソフトウェアである。本研究では、そのうちの 1 つである全文検索システム Namazu (以下 Namazu) を取り扱っている。

Namazu とは、手軽に使えることを第一目標にした日本語全文検索システムであり、作者は愛知大学の高林哲氏である。CGI として動作させることにより、小中規模の WWW 全文検索システムを構築することができるほか、ハードディスク内のファイルを対象としたようなパーソナルな用途にも使えるようになっている。Namazu には、フリーソフトとしては画期的な以下のような多くの特徴がある。

- (1) 各種プラットフォームに対応
- (2) 高速な検索を実現
- (3) 多くの検索方法をサポート
- (4) 検索結果表示の柔軟性
- (5) 正確な HTML の取扱い
- (6) 複数インデックスからの検索をサポート
- (7) 多彩な検索クライアント

これらの特徴を生かし、NEC エンジニアリングの広瀬健一氏が移植した Microsoft Windows 98 上で動作するバージョンのもの、2 つの利用方法を検討した。

はじめに、イントラネット内専用のサーチエンジンを構築した。これにより、イントラネット上のファイルの内容での検索が可能になり、ファイルの所在が不明になることが非常に少なくなる。本研究では、本大学建設情報研究室内のイントラネットに Namazu サーバを設置し、研究室内の複数台の PC に対してそれぞれにインデックスを作成して、複数インデックスからの検索の動作確認をした。サーバには、A. Nakata 氏が製作した「AN httpd」と Microsoft 社の「Personal Web Server」を用いたが、どちらも正常に動作した。図 1 に検索結果画面の例を示す。

次に、後述する Microsoft Access 2000 (以下 Access) で作成した工事文章リンクの補助的機能としての利用を検討した。Access のようなリレーショナルデータベース管理システム (以下 RDBMS) の場合、1 万件や 10 万件、100 万件といった多量のデータレコードに対して検索を行うと、検索結果を得るまでの反応速度が極端に遅くなる。それに対し、全文検索システムは、検索対象のデータの多少に関わらず RDBMS と比較して非常に高速であり、また検索結果を得るまでの反応速度は、文章数の対数 (文章数を n とすれば、 $\log(n)$) に比例する。つまり、この Namazu の利用方法は、巨大データベースの検索における検索時間の爆発の回避策に相当する。

Namazu の問題点として、XML (eXtensible Markup Language) に対応していないことや、フリーのソフトウェアであるため、漢字コードの変換と HTML の解析以外は何も行わず、そのため一般的なワープロソフトのフォーマットに対応していないことなどが挙げられる。

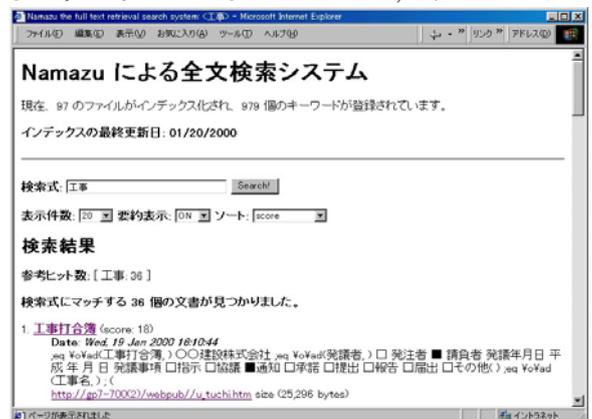


図 1. 全文検索システム Namazu の検索結果画面

4. Microsoft Access2000

AccessはWindowsプラットフォームにおいて最も代表的なRDBMSである³⁾。その利用方法は多岐にわたり、特に1999年7月に発売されたAccess2000では、従来のAccessと比べWeb機能が大幅に強化されているので、その使用範囲をインターネットに対し大きく広げている。本研究ではAccessを用いて2つのアプリケーションを作成し、それらを通じてAccessの建設CALs/EC推進のための利用方法を模索した。

1つは、工事文章のデータベースへの検索をAccessで行うものである(図2)。建設CALs/ECで用いられると考えられるCALs標準文章は、40種類以上にのぼり、一つの工事に対して膨大な文章が作成されると考えられる。このとき作られる文章のデータベースに対し、ファイル名だけで検索を行うのは、たとえファイル名が特徴的なものであってもわかりにくい。さらに、ファイル名に載せることができる情報はごく限られているため、それだけの検索では多くの検索要求に応えることはできない。このアプリケーションでは、文章のファイル名とそのリンク、文章が使われた工事の名前、文章名、記入日を1つのデータレコードとしてデータベースを構築することで、これら複数の条件からリンクを検索し、目的の文章を探し出すことを可能にしている。Accessを含むRDBMSの欠点として、多量のデータレコードに対し、検索の反応速度が遅くなることがあげられるが、文章の内容そのものをデータベース化するのではなく、文章の特徴とリンクのみをデータベース化することと、Namazuと併用することで、その問題を回避している。

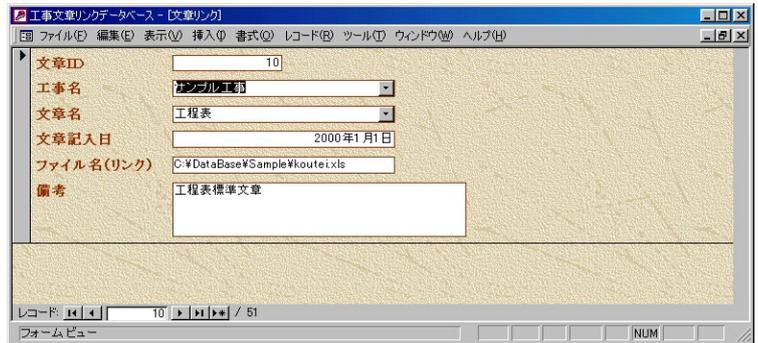


図 2. 工事文章リンクのフォーム



図 3. ネットワーク式工程表のフォーム

他の1つは、ネットワーク式工程表をもとに、毎日の施工状況の記録をAccessで行うものである(図3)。建設CALs/ECにおいて、データの電子化は必要不可欠であり、紙の状態データを保存し、そのデータを電子化するよりも、直接データを電子的に保存する方が作業時間短縮につながり望ましい。さらに、建設省の建設CALs/EC実証フィールド実験⁴⁾において、工事管理情報のデータ電子化作業が他の作業に比べて時間がかかっていることから、建設CALs/ECの推進には、データの電子化作業の簡易化が必要であると考えられる。このアプリケーションは、工事名や日付に対応した項目の自動入力や、コンボボックスなどを用いて、利用者の入力作業を簡易化できるように設計されている。ただし、実際の業務でこのアプリケーションを利用するには、より多くの内容を取り入れ、それぞれの関連性を深めなければならない。

両アプリケーションとも、インターネットでの利用を想定し、静的なHTMLファイル、Accessの最新データを基に動的にHTMLファイルが生成されるアクティブサーバページ、インターネット経由でAccessと交信してデータを表示及び操作できるデータアクセスページの3種類での動作を確認したが、いずれも自由度が低く、それぞれの長所と短所を生かした利用法を考える必要がある。

5. おわりに

建設CALs/ECでは、巨大な統合データベースの構築が予想される。インターネットにも見られるように、データベースが巨大になればなるほど気軽に操作できる情報検索システムの需要が出るはずであり、その問題の答えの1つがNamazuのような高速な情報検索システムである。建設CALs/ECでは文章以外のCADなどのデータもデータベースに含まれるので、今後それらに対応する検索システムの開発が求められる。さらに、本研究のAccessで作成したアプリケーションを現時点で実利用することはできないが、これらのようなデータ入力の簡易化を図るアプリケーションは建設CALs/ECの推進に必要であると考えられる。建設CALs/ECの統合データベースを効率よく利用するためには、それ自体の大きな問題点である標準フォーマットの解決と、そのフォーマットに対応できる情報検索システムの開発が必要であると考えられる。

【参考文献】

- 1) 馬場肇:日本語全文検索システムの構築と活用,ソフトバンク株式会社,1998.9.25.
- 2) 土木情報システム委員会・情報共有技術小委員会:平成9年度活動報告書,社団法人土木学会,pp.11-21,1998.6.1.
- 3) 高羽実:はじめてのAccess2000,株式会社秀和システム,1999.7.25.
- 4) 関西大学総合情報学部・建設CALs 関西研究調査協議会:
建設分野におけるCAD/CG/GIS/GPSの関西情報発信セミナー講演資料,pp.8-13,1999.12.1.