



情報検索システムの土木 分野での利用方法の研究

担当教員: 皆川 勝

学生氏名: 永井達也



はじめに

- ◆ 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変化を遂げた。
- ◆ 情報検索システムは建設省が進める建設CALS/ECの普及及び管理に大きな影響を与えると考えられる。
- ◆ 本研究では、情報検索システムの土木分野での有効利用法を模索することを目的とする。



建設CALS/EC

- ◆ 米軍後方補給支援に使われた技術と、電子商取引(EC)を組み合わせ、建設分野での利用を目的とした活動及び概念。
- ◆ 紙の資料を電子化し、その情報を複数の役所や企業外の会社で共有する。
- ◆ 設備機器・ソフト他多くの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。



全文検索システムNamazu

- ◆ インターネットの加速的な進展に伴い、情報の所得が困難になった。
- ◆ この状態を回避すべく開発されたのが全文検索システムである。
- ◆ 本研究では全文検索システムNamazuの、土木分野での利用方法を模索した。



Namazuの研究内容

- ◆ 複数インデックスによる検索
- ◆ XMLへの対応の照査
- ◆ リレーショナルデータベース管理システム (RDBMS)の補助

複数インデックスによる検索



C:\Nagai\Namazu\search.html - Microsoft Internet Explorer

ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(T) リンク アドレス(D)

建設情報研究室 検索ページ

以下のフォームから当サイトのWebページの内容を検索することができます。
検索式欄にキーワードを入力し、Search!ボタンを押してください。

検索式: Search!

表示件数: 20

対象インデックス

- Aptivablack CDドライブ
- Aptivablack Dドライブ
- Aptiva white CDドライブ
- Aptiva white Dドライブ
- Aptiva white Eドライブ
- Fmvsll-167 CDドライブ
- Fmvsll-167 Eドライブ
- Gatewaygp450 CDドライブ
- Gatewaygp450 Dドライブ

ページが表示されました

マイコンピュータ

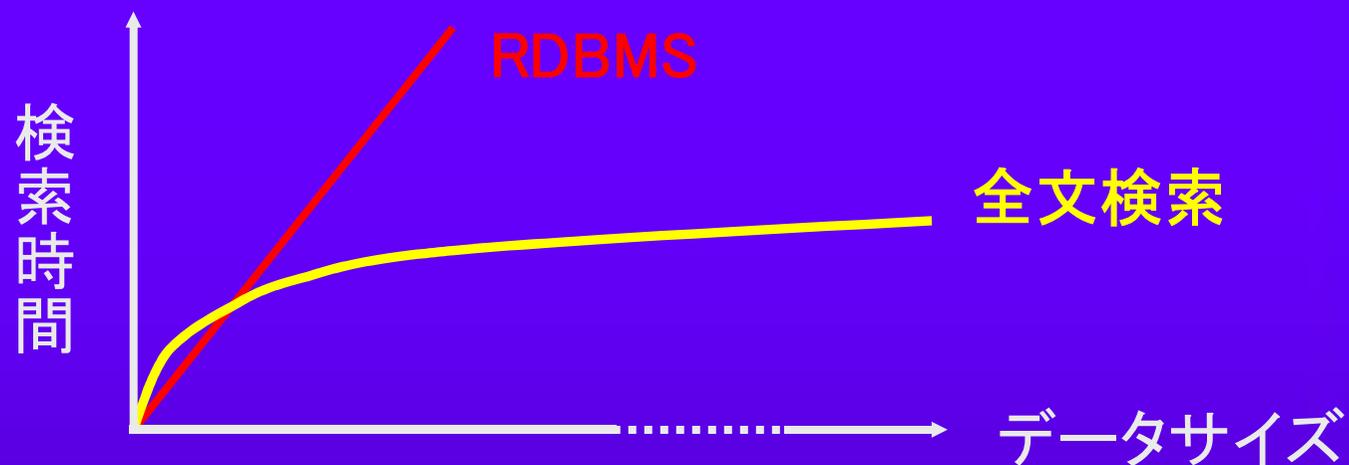
チェックボタンによる
インデックスの選択



XML対応の照査

- ◆ テキストファイル、MS-Word2000で作成したドキュメントファイル及びHTMLファイル、XMLファイルを検索対象とし、インデックスの作成及び検索に関する比較を行った。
- ◆ 結果、NamazuはXMLに対応していないことがわかった。

RDBMSの補助



- ◆ RDBMSは多量のデータに対し検索を行うと結果を得るまでの反応速度が極端に遅くなるという欠点をもつ。
- ◆ 全文検索システムの反応速度は文章数の対数に比例する。



Namazuにおける結論

- ◆ Namazuは文章を管理するデータベースにおいて効果を発揮することから、土木分野においても公文書など文章データベースの基幹エンジンとしての使用法が基本的かつ最も効果的と考えられる。
- ◆ 巨大データベースに対しても効果を発揮するが、XMLに対応していないので、今後の開発が求められる。



Microsoft Access2000

- ◆ データベースとは、様々な業務において操作の対象となるデータの集合を意味し、そのデータベースを管理するソフトウェアをデータベース管理システム(DBMS)という。
- ◆ 本研究では、Windowsプラットフォームにおいて最も代表的なRDBMSであるMicrosoft Access2000の、土木分野での利用方法を模索した。



Accessの研究内容

- ◆ 2つのアプリケーションを作成し、それらを通じて利用方法の模索を行った。作成したアプリケーションは以下の2つである。
 - 工事文章リンク
 - ネットワーク式工程表
- ◆ さらに、従来のAccessと比較し大幅な強化がなされたWeb機能でのアプリケーションの利用も照査した。

工事文章リンク

- ◆ 建設CALS/ECで用いられると考えられる多量の文章によるデータベースに対し、複数の条件から検索できるアプリケーションを作成した。



文章ID	10
工事名	サンプル工事
文章名	工程表
文章記入日	2000年1月1日
ファイル名(リンク)	C:¥DataBase¥Sample¥koutei.xls
備考	工程表標準文章

レコード: 1 / 1 (フィルタ)

ネットワーク式工程表

- ◆ 建設CALS/ECではデータ電子化作業が必須であるため、日毎の工事管理データの電子化支援アプリケーションを作成した。

作業状況 ネットワーク式工程表

作業番号	作業名	クリティカルパス	予定工程%	実施工程%	備考
4 B			100.0%		
5 E			40.0%		
6 F			60.0%		
7 G		最重要管理経路	6.7%		

レコード: 1 / 4

レコード: 36 / 80



アプリケーションのWeb利用

- ◆ Accessの機能を使って2つのアプリケーションをWeb上で試行した。
- ◆ 利用は可能だが、自由度及びUIが著しく低下することがわかった。

Namazu による全文検索システム
現在、97 のファイルがインデックス化され、979 個のキーワードが登録されています。
インデックスの最終更新日: 01/20/2000

検索式: Search!

表示件数: 20 要約表示: ON ソート: score

検索の方法
大文字、小文字の区別はありません。foo*のように末尾にアスタリッシュで前方一致検索が可能です。また、単語をスペース区切りで並べて書きます。日本語は KAKASI/ChaSen によって分解され、「日本語情報本語」「情報処理」のように2つの単語に分かれてアンド検索されます。語の分解は完全ではありません。品質は辞書によって決定されます。

ページが表示されました

700(2)サーバーテストページ
ようこそ
今日は 00/02/17 です。

工事文章リンク

文章ID	1
工事名	サンプル工事
工事名ふりがな	さんぷるこうじ
工事着工年月日	2000年1月1日
工事完成年月日	2000年12月31日
文章分類名	調査系
文章名	打合せ記録簿
文章記入日	2000年1月1日
ファイル名(リンク)	http://gp7-700(2)/webpub/utiawase.htm
備考	打合せ記録簿標準文章

工程管理入力テーブル

ID:

作業名:

実施工程%:

備考:

管理入力テーブル

データアクセスページによる工程状況確認クエリへ
アクティブサーバーページによる工程状況確認クエリへ



Accessにおける結論

- ◆ Accessは関連性を持ったデータ群の管理に効果を発揮する汎用性の高いDBMSである。
- ◆ アプリケーションの開発によって用途の広い利用が可能である。
- ◆ インターネットで利用する場合UIが大きく損なわれるので、現時点での建設CALS/ECでの利用は難しいと考えられる。



おわりに(1)

- ◆ インターネットに見られるように、情報検索システムの必要性は非常に高い。
- ◆ 建設CALS/ECでは、多種多様なデータを扱うデータベースの構築が予想されることから、オブジェクト指向のDBMSが必要であり、その開発が求められる。



おわりに(2)

- ◆ 建設CALS/ECの課題の1つである標準フォーマットについて、データベースの基幹エンジンとなるDBMSにあわせてその規格を定めるという考え方もある。
- ◆ この場合の問題点として、
 - 自由度が少ない
 - 入力支援ソフトの開発と普及などが考えられる。