

Webコンピューティング技術について

山梨大学
大学院 医学工学総合研究部
渡辺喜道

目次

- Web検索の効率化
 - 類義語辞典の利用
 - スニペットの利用
- Webメールのインタフェースの改善

2006/5/20

Web検索における課題

- Yahoo!やInfoSeekなどの検索サービスでキーワード検索すると、期待した内容とは関係のないページまでが数多く出てきて必要な情報を探し出すのに一苦労、なんて経験は誰にでもあることだろう。
- 何とかしたい。



2006/5/20

サーチエンジンの分類

- カテゴリ系(Yahooなど)
 - 各ページに関する簡単な説明を検索対象とするもの。ページはカテゴリ毎に整理されている。
 - 人が確認してから登録されるため、良質なサイトが多い。
- ロボット系(Googleなど)
 - ロボットで収集したデータから(全文)検索インデックスを作成して提供するシステム。
 - ロボットプログラムが巡回してデータを収集するため、情報量が多い。
- メタサーチ
 - 一度に複数の検索エンジンに問い合わせ、結果を統合して表示するシステム。

2006/5/20

検索支援ツールZoom (Ask Jeeves)

- ユーザが入力した検索キーワードから検索結果を絞り込む(=zoom in)または、関連する検索対象を広げる(=zoom out)機能を持つ。
- Zoomは視覚的に表現され、画面右側にカテゴリ化されて表示される。
- 検索技術Teoma独自のクラスタリング技術を使っている。

2006/5/20

Zoomのありがたみ

- 3つのセクションに分けて、複数の検索キーワード候補を表示する。
 - 現在の検索結果をさらに絞り込む、またはより明確化するキーワード候補を表示する。
 - 現在の検索キーワードから連想しうるキーワード候補や、より広範囲なキーワード候補を表示する。
 - 関連する人物名を検索キーワード候補として表示する。

2006/5/20

検索エンジンTeoma

7

- データベースの中にあるWebサイトをその内容に応じてどれだけ多くの“権威あるサイト”からリンクされているかという情報に基づいて仕分けする。
- さらに、同じ内容を扱っているサイトの“コミュニティ”を構築し、その中からその分野のエキスパートを発見するという方法をとっている。
- コミュニティと検索語を対応付けている。

2006/5/20

検索支援ツールWeb Answers (Ask Jeeves)

8

- 検索エンジンに登録している数十億のWebページから、質問に対する回答をリアルタイムで検索するツール
- 例えば、“Deadliest snake” (最も危険なヘビ)、“what is data mining” (データマイニングとは?)といった質問に対する回答を探し出して表示する。

2006/5/20

情報共有型検索支援ツール

9

- ソーシャルブックマーク
 - 自分がブックマークしているサイトのリストをネット上に公開し、他のユーザと共有することができるサービス。
 - 自分自身の情報管理や話題の情報のチェックに利用することができる。
 - 多くのソーシャルブックマークには、関連する話題を芋づる式にピックアップしていくことができる。
- 検索バディ
 - 同じキーワードで検索した人にオススメのサイトを教えてもらう。

2006/5/20

研究室での取り組み

10

- 効率よく情報検索するために
 - 1 キーワードの類義語や関連語を利用する。
 - 2 SERP (Search Engine Result Page)情報を利用する。

2006/5/20

類義語・関連語の利用

11

- どんなキーワードで検索したらよいかわからない。
- 必要な情報が得られるキーワードが思い付かない。
- Google検索サイト
 - “~”を付けたキーワードについて、自動で類義語によるWebページ検索を実行する。
 - しかし、英単語にしか対応していない。
 - また、ユーザが意図しない検索キーワードによる検索が行われる可能性がある。

2006/5/20

類義語を用いた検索

12

- 「デジタル類語辞典 第3版 シソーラス」を利用して、類義語が提示される。
- ユーザは利用したい類義語を選択し、検索をかける。

2006/5/20

ツールの実行例(1)

13

類義語Web検索システム

類義語検索を行うキーワードを入力して下さい

キーワード入力

**類義語検索を行いたい
キーワードを入力**

2006/5/20

ツールの実行例(2)

14

類義語検索結果

計算機の類義語として以下のものが見つかりました。
Webページ検索に使用する類義語を指定して下さい。

“計算機”で類義語検索した例

- 計算機
- コンピュータ
- コンピューター
- パソコン
- 電算
- 電算機
- 電子計算機
- 電話

Webページ検索に使用するキーワードのチェックボックスをチェック

2006/5/20

ツールの実行例(3)

15

類義語検索結果

計算機の類義語として以下のものが見つかりました。
Webページ検索に使用する類義語を指定して下さい。

- 計算機
- コンピュータ
- コンピューター
- パソコン
- 電算
- 電算機
- 電子計算機
- 電話

2006/5/20

ツールの実行例(4)

16

検索オプション指定

類義語以外の検索条件を指定して下さい。

OR:

AND:

NOT:

2006/5/20

検索結果の利用

17

検索結果ページからホームページをクリック(訪問)する基準は何ですか？

基準	割合
タイトルと説明文を総合的に判断	64%
上から順番にクリック	17%
タイトルを読んで	16%
説明文を読んで	12%
適切に	3%

http://www.semseminar.com/anq_search.html

2006/5/20

スニペット

18

2006/5/20

スニペットの構造

19

- 文全体の長さは100文字程度である。
- 一文または複数の文で構成される。
 - 複数の文で構成される場合は“...”で文を区切る。
- 検索エンジンが対象Webページから文を抽出し、スニペットを作る。

2006/5/20

スニペットを利用した検索の効率化

20

- スニペットの情報からそのページにジャンプする。
- ジャンプ後は、ページの先頭から表示されてしまう。
- ジャンプしたとき、スニペットで表示されていた部分を見たい。
- <A>タグを追加し、目的の箇所にジャンプする。

2006/5/20

スニペットの表示方法の違い

21

- Googleの特徴
 - メタタグに検索キーワードが含まれる場合のみ、メタタグの内容がスニペットに表示される。
 - キーワードを含む文は30文字以上。
- Yahoo!の特徴
 - メタタグが存在すれば、その内容が必ずスニペットに表示される。
 - キーワードを含む文は20文字以上。

2006/5/20

ツールの操作手順

22



2006/5/20

ツールの操作手順

23



2006/5/20

ツールの操作手順

24



2006/5/20

ツールの評価

25

- 利点
 - 求める情報であるかどうかの判断が容易になった。
 - 既存の手順では、7.98秒。
 - ツールを使うと、2.01秒。
- 欠点
 - プラグインになっていない。
 - スニペット文が必ず最上部に表示されるわけではない。
 - 既存のブラウザと比較すると付加機能が足りない。

2006/5/20

Webメールの利用

26

- インターネット上でのコミュニケーション手段として電子メールを用いることが多い。
- Webメールはどこからでもメールが読めるという特徴があるが、メーラよりもインタフェースや機能が必ずしも十分ではない。

2006/5/20

既存のWebメールの課題

27



- 動作が遅い。
- ユーザインタフェースが貧しい。

2006/5/20

Webメールの改良点

28

- メールをドラッグ&ドロップで移動できる。
- ポップアップメニューから操作できる。
- メールをプレビューできる。
- 添付ファイルをドラッグ&ドロップで操作できる。
- メールアドレスのオートコンプリート機能をつける。
- メールプレビュー内容、メール作成画面をあらかじめロードしておくことができる。
- フォルダ切り替え時にページ全体をロードせず、メール一覧データのみをロードする。

2006/5/20

利用した技術

29

- JavaScript
- CSS
- Ajax
- DOM
- DHTML
- Cgi
- Perl

2006/5/20

実行例

30



2006/5/20

評価

31

作業時間	既存	ツール
メールをゴミ箱に移動する	1.80秒	0.40秒
メール10件をゴミ箱に移動する	4.35秒	4.10秒
添付ファイルを追加する	7.75秒	2.10秒
添付ファイルを10個追加する	60.00秒	2.30秒
メール10件の概要を把握する	25.30秒	4.30秒
送信画面の表示する	0.85秒	0.10秒
フォルダを切り替える	0.85秒	0.75秒

2006/5/20

研究室の取り組み

32

- 半構造文書(XML文書)処理系
 - 意味的に整合性の取れた文書作成支援
 - アクセス制御による文書管理
 - データ変更に頑健なデータベース型文書管理
 - インターネット上の表形式データの統合管理
 - Webアクセシビリティの向上
 - リンク切れのない文書管理

2006/5/20