

データベースと情報検索

情報検索(1)

検索エンジンを使ってみる

工学部担当

教員 岩村 雅一

日程(情報検索:担当 岩村)

- 12/9 検索エンジンを使ってみる
 - 12/16 メディア検索を使ってみる
 - 12/25 ウェブアプリケーションを使ってみる
 - 1/9 検索エンジンを用いた演習
 - 1/20 検索エンジンの仕組み
 - 1/27 メディア検索の仕組み
 - 2/3 消費者生成メディアの最近
-

情報検索 (information retrieval)

- 予め蓄積された情報の中から特定の目的にあった情報を探し出すこと
 - メディアによる分類
 - テキスト、Web, 画像、ビデオ、音楽、地図、、、
 - ここでは当面Webの検索を主に考える
 - Web検索に求められること
 - 高い適合率
-

適合率と再現率

検索された
ページ

正解の
ページ

A

B

適合率: $\frac{|A \cap B|}{|A|}$

再現率: $\frac{|A \cap B|}{|B|}$

適合率と再現率

- Web検索はなぜ適合率重視？
 - 再現率を重視する検索には何がある？
-

本日のメニュー

- 様々な検索エンジンの紹介
 - Googleの色々な使い方
-

代表的な検索サイト

- Web検索
 - www.google.co.jp
 - 本の検索
 - Google books: books.google.co.jp
 - 国立国会図書館 www.ndl.go.jp
 - amazon.co.jp, [.com](http://amazon.com)
 - Webcat Plus(図書検索) webcatplus.nii.ac.jp
 - 論文検索
 - citeseerx.ist.psu.edu
 - scholar.google.com
 - 特許検索
 - www.ipdl.inpit.go.jp/homepg.ipdl
 - 総合
 - 例えば, www.searchdesk.com
-

実習 googleを使って

- キーワードを入れるだけではない！
- 参考文献
 - Google Hacks , オライリージャパン

AND, OR

□ AND

- ならべるだけ(いつもの方法)

□ OR

- 「 $x \text{ OR } y$ 」, 「 $x \mid y$ 」と書く; “または”の意味
- 「 $A (B \mid C)$ 」はどのような意味?
- 例: 橋下徹 (大阪府立大学 | やしきたかじん)

□ NOT

マイナス検索、プラス検索

□ マイナス検索

- notの意味
- キーワードxを含まない -x
- 「阪神 -タイガース」

□ プラス検索

- ストップワード(a the など)を含める
 - 「+the king」
-

フレーズ検索

- “radio button”
 - and, or, notも使える
 - snowblower snowmobile –”green bay”
-

類義語検索

□ ~をつける

- 「~ape」とすると、「monkey, gorilla」などにも
マッチ
-

数値の範囲指定

- 「x..y」でx以上y以下を表す
 - 「numrange:x-y」でもよい
 - 例:「デジタルカメラ 900..1200 万画素」
-

ワイルドカード

- 「three * mice」 *は任意の1単語に対応
 - 「three blind mice」、「three blue mice」などにマッチする
 - 「*の都」は「杜の都」、「水の都」にマッチ
 - 英語の用法を調べたいとき、歌詞を調べたいときなどに役立つ
-

32語の制限

- 検索質問として与えられるのは32語まで
 - ワイルドカードを使う場合を除く
 - “It has been three years since Osaka Prefecture University, Osaka Women's University and Osaka Prefecture College of Nursing integrated into one university, incorporated and reemerged as the Osaka Prefecture”
なら、どのページ？
-

特別構文(1)

- intitle: タイトルのみを検索
 - 「intitle:"george bush"」
 - 「allintitle:"money supply" economics」
 - intext: 本文のみを検索
 - 「intext:"yahoo.com"」
 - inanchor: アンカーテキストを検索
 - 「inanchor:"tom peters"」
 - site: サイトを限定
 - 「site:osakafu-u.ac.jp」
-

特別構文(2)

- inurl: URLの中にある語を検索対象とする
 - 「inurl:help」
 - 「allinurl:search help」
 - link: 指定したURLにリンクしているページを検索
 - 「link:www.osakafu-u.ac.jp」
 - daterange: 日付範囲指定検索
 - 「daterange:開始日ー終了日」
 - 日の指定にはユリウス日を使う
 - 「"spice girls" daterange:2450958-2450968」
-

特別構文(3)

- filetype: 拡張子の検索
 - 「filetype:ppt google」
 - related: 関連ページの検索
 - 「related:yahoo.com」
-

特別構文の利用

□ 検索オプションのページ

- http://www.google.co.jp/advanced_search?hl=ja

その他のTIPS

□ define: 専門用語などの定義

■ define:RFID

様々なページ・サイト

□ トренд分析 Zeitgeist

- <http://www.google.com/intl/en/press/zeitgeist/>

□ Googleには他にも様々なサービスあり

- <http://www.google.co.jp/intl/ja/options/>

□ 人物検索

- <http://spysee.jp/>
-

研究レベルの検索サイト

□ goo ラボ

- <http://labs.goo.ne.jp/>
 - 色々面白い検索エンジンあり
-

レポート課題

□ 面白い検索例を報告する

■ 面白い検索例

- 皆があまり知らないような検索方法
- 皆があまり知らないようなサイト
- 皆があまり知らないようなコンテンツ
- 探し出すのが簡単ではないページ
 - 検索質問に工夫があるもの

□ 公序良俗に反するものは却下

□ レポートにまとめて、次週授業開始時に提出

レポートの内容

- 検索の目的
 - 検索の方法
 - サイトや質問をどうしたか？
 - 検索の結果
 - 何が出てきて、どう面白かったのか
 - 分量： **A4で1枚以上**
 - ~~□ ファイル形式：PDF、Word、テキストファイルなど~~
-

レポート提出方法

- ~~授業支援システムに行き、「データベースと情報検索」を選択~~
 - ~~第9回の「課題：検索エンジンを使ってみる」を選択し、レポートをアップロード~~
 - ~~×切：次回授業開始時まで~~
 - ~~×切後でもアップロードできるが、採点しない~~
-