

情報検索演習

第4回資料

-Webページ、Webサイトの探し方(1)-

鶴見大学 非常勤講師
江草由佳
(国立教育政策研究所 総括研究官)
yuka@nier.go.jp

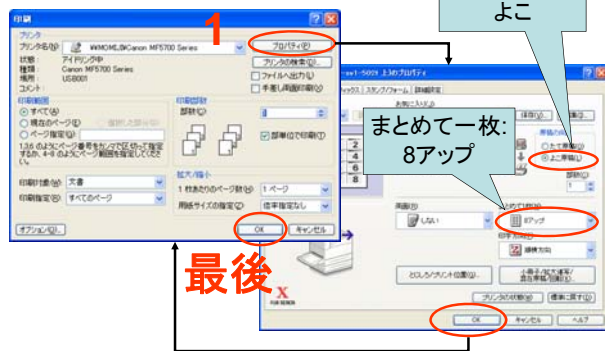
本日のお品書き

- 演習: Webページ、Webサイトの探し方
 - 検索エンジン (Google, Yahoo! Japan)
 - メタ検索エンジン (検索デスク, Ceek.jp)
- 第4回演習課題

第4回演習課題

- 演習課題を実行し、講義Webサイトから、回答用紙をダウンロードし、PowerPointで作成して、8分割印刷して提出しなさい。
 - 課題名に: 第__回演習課題と入れなさい
 - 〆切: 来週演習開始時
- 「情報検索演習」、「第__回演習課題・Webページ、Webサイトの探し方(1)」、「時限」「提出日」「学籍番号」、「名前」、を忘れずに記入すること

参考: 8ページをまとめて一枚として印刷する方法



Webでの情報の単位

- Webページ
 - Webページブラウザで一度に表示できるひとまとまりの情報
- Webサイト
 - Webページをひとまとまりに公開するWebページ群
- ホームページ、トップページ
 - Webサイトの最初のページ
 - Webページのことをさしてホームページと呼ぶことが多いが正確な表現ではない

Webページ、Webサイトの特徴とそのアプローチ

- 網羅的なもの
 - 検索エンジン
 - Webページ、Webサイトを網羅的に収集して検索できる
 - アーカイブサイト
 - 過去のある時点でのWeb上の情報を閲覧できる
- 特定の範囲や分野に限定
 - サイト内検索
 - サブジェクトゲートウェイ、分野別リンク集
 - 画像検索サイト
 - ...

検索エンジン(1)

- サーチエンジンとも言う
- Webページを網羅的に収集し、検索可能にしたサービス
- 種類
 - ロボット型 (全文検索型)
 - ディレクトリ型 (カテゴリ型)
 - メタ検索エンジン
- ポータルサイトの核となる一機能

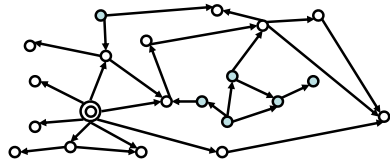
検索エンジン(2)



ロボット型(全文型)検索エンジン(1)

- 代表的なもの: Google
- クローラ(収集ロボット)とよばれる収集プログラムでWebページを網羅的に収集
- 登録しているサイトは膨大
- ディレクトリ型に比べて
 - 更新頻度が高い
 - 検索結果の精度が低くなりがち
- ランキング手法
 - 関連(スコア)の高いものから順に表示する

Webの模式図



10

ロボット型(全文型)検索エンジン(2)

- 注意1: 全てを収集しているわけではない
 - 例: robot.txt で収集拒否を宣言
 - 例: アクセスにIDとパスワードが必要
- 注意2: 検索は収集した時点でのWebページ
 - 現在閲覧できるWebページとは異なることがある
 - 収集した時点でのページを見れる機能がある
 - 一般にキャッシュ機能と呼ばれる

11

Google(1): <http://www.google.co.jp/>

- リンク情報を利用したランキングのパイオニア
- Google: 検索式(演算子は全て半角)
 - AND検索: `␣`(半角スペース) 例: 英語 多読
 - OR検索: `OR` 例: 高校 OR 高等学校
 - Not検索: `-` 例: 車 -トラック
 - フレーズ検索: `" "`でくる 例: "図書館と著作権"
- 検索オプションでクエリボックスやメニューをつかって複雑な検索可能
- ログインして検索→個人に適した検索結果

12

Google(2)

- p.44【例題1】
 - 久留米出身の洋画家で坂本繁二郎を除く、青木繁か古賀春江のサイトを検索したい
 - 検索クエリボックスに次のように入力して検索
 - ※ `␣ ←`は半角空白をあらわす
- 古賀春江␣OR␣青木繁␣-坂本繁二郎
- 検索結果が表示される→画面イメージをpptへ

13

Google 検索オプション使用例



14 (実行 2012-10-05)

ディレクトリ型検索エンジン

- 代表的なもの: Yahoo! カテゴリ
- カテゴリ型検索エンジンともいう
- Webサイトをカテゴリごとに分類してたどれるようにしたもの
- 求める主題やトピックをたどってブラウジングができる
- 人間が:
 - Webサイトを一定の収集基準の元に採択可否を決定
 - Webサイトをカテゴリに仕分け
 - ページの説明情報を作成
- ロボット型に比べて
 - 登録されているサイトが少ない
 - ノイズが少ない
 - 登録までに時間がかかる

Webサイトが膨大になってきて人手を介して更新するにはコストが膨大なため、最近では提供するところが少なくなってきた。

15

Yahoo! カテゴリ(1) <http://dir.yahoo.co.jp/>

- Yahoo! JAPANの一サービス
 - Yahoo!JAPAN: 日本でよく利用されているポータルサイト(<http://www.yahoo.co.jp/>)
- 日本人に有用と考えられるWebサイトを収集
- 簡単な解説をサイトごとに付与
- 主題ごとに階層的になっている
- アナリストは200人といわれている

16

Yahoo! カテゴリ(2)

- P.45【例題2】リンクをたどって「教育図書館」をたどってみましょう

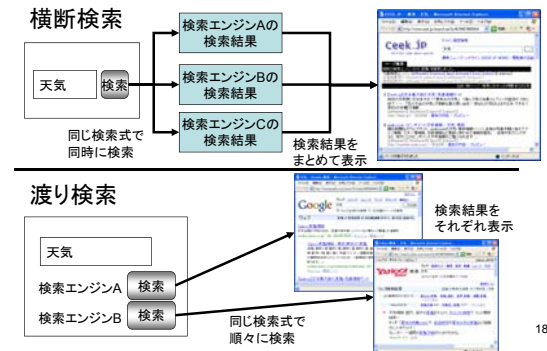


- ↑の画面イメージをpptへ

17

メタ検索エンジン(1)

複数の検索エンジンを同時に/並行的に検索できる検索エンジン



18

メタ検索エンジン(2)

- Ceek.jp
 - <http://www.ceek.jp/>
 - 横断検索型メタ検索エンジン
 - 複数の検索エンジンを同時検索
 - 検索結果を統合表示
- 検索デスク SearchDesk
 - <http://www.searchdesk.com/>
 - 渡り検索型メタ検索エンジン
 - キーワード入力はそのままで次々と異なった検索エンジンへ
 - よく利用される検索エンジンはより上位に表示
 - 辞書サイトなども検索できる

19

メタ検索エンジン(3)

- P.46【例題3】読書習慣について調べたい。
いろいろな検索エンジンの検索結果をまとめて得たい。
 - Ceek.jpを検索した結果→画面イメージをpptへ
- P.46【例題4】読書習慣について調べたい。
いろいろな検索エンジンの検索結果をそれぞれ見比べたい。
 - 検索デスクを検索した結果→画面イメージ(検索デスクとGoogleとbingを表示した画面全体)をpptへ

20

ポータルサイト

- ポータル(portal): 入り口
- このサイトを入り口=スタート地点にしてそこからいろいろなサイトへ行くためのサイト
- 普段よく使う様々なサービスがある
 - Webメールサービス、ニュース、路線検索、地図検索、辞書検索...
- 検索エンジンのみを提供していたサービス(Google, Yahoo!など)はポータルサイトへと移行していった
 - 検索エンジンはポータルサイトの核となるサービスの一つ
- 検索ののっかかりとして利用できる

21

検索エンジンを使うときの 一般的な注意

1. Webページ、Webサイトが検索できるわけではない
2. 検索エンジンによって結果が大きく異なる
3. 検索結果の質や信頼性を確認する
4. ヘルプを活用する
5. キーワードはWebページに表記されたものが対象となる
6. 検索式の入力には英数字、記号は半角と全角では異なる挙動をする場合があるので気を付けて使う

22