

情報検索演習

第5回資料

Webページ、Webサイトの探し方(2)

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

- アーカイブサイト(WayBackMachin, WARP)
- サイト内検索
- サイトマップ
- 特定のメディアに限定した検索
- 第5回演習課題: Webページ、Webサイトの探し方(2)
- 第6回演習課題: サーチエンジンの比較

第5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、9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
 - ✕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(2)」、「時限」「提出日」「学籍番号」、「名前」、を忘れずに記入すること

アーカイブ・サイト

- 定期的にWeb上の情報資源を保存・提供
- 過去のWebページが見れる
 - 検索エンジン経由では、現在のWebページもしくは収集時点のもの(キャッシュ)しかみれない
 - 消えてしまったWebページ(例:「404 Not Found Error」)、更新されて変更になった過去のWebページなどが見れる
- 検索可能(URLやキーワードなど)
- 代表的なものに、Wayback Machine とインターネット資料収集保存事業(WARP)がある

Wayback Machine

- <http://archive.org/web/web.php>
- 米国非営利団体 Internet Archive が運営
- 巨大なアーカイブ・サイト
- 世界の多くのサイトを収集
 - 収集していないページもある
 - (例: robot.txtで収集拒否のWebページは集めていない)
- 1996年以降から収集開始
- URLを指定して過去のWebページが閲覧できる

インターネット資料収集保存事業 (WARP)

- <http://warp.ndl.go.jp/>
- 国立国会図書館が運営
- ウェブ・アーカイブ事業
- 日本の中央省庁、立法機関、都道府県、各種法人、国立大学のWebページなど
- Wayback Machineよりも収集頻度は低い
- 2002年から収集開始
- キーワードや編者を指定した検索ができる
- テーマや分類からたどれる
- 本文の全文が検索できる

演習

- 日本首相官邸の過去のWebページをアーカイブ・サイトで見てみよう。
 - 首相官邸ホームページ:
<http://www.kantei.go.jp/>
- 1. その前に現在の首相官邸ページをみてみよう
 - 1. IEを起動
 - 2. <http://www.kantei.go.jp/> を入力し「Enter」
→画面イメージをpptへ

演習

— 2. WayBackMachine —

1. WayBackMachineのサイトに行く
 - <http://archive.org/web/web.php>
 2. 首相官邸のホームページのURLを入力
 - <http://www.kantei.go.jp/>
 3. 「Take Me Back」ボタンをクリック
 4. カレンダーの日付のリンクをクリックして過去のWebページをみる
 - 1つだけではなく、いろいろみってみる！
 - 例：1997年の首相は誰だったかな？
 - 例：2000年は？
- どれか一つの日付の画面イメージをpptへ

演習

— 3. WARP—

1. WARPのサイトに行く
 - <http://warp.ndl.go.jp/>
2. キーワード:「首相官邸」を入力
3. 「検索」ボタンをクリック
4. 「首相官邸／内閣官房」のリンクをクリックして過去のWebページをみる
5. 「2004年11月19日」収集 のリンクをクリック →画面イメージをpptへ

特定のWebサイト内の検索(1)

--サイト内検索--

- Webサイト内のみを検索する
 - 検索エンジンのドメイン指定などでも代用できるが、、、
- Webサイトに独自の検索機能やインターフェースがあることも
- 例 : Sony
 - <http://www.sony.co.jp/Search/>

特定のWebサイト内の検索(2)

--サイトマップ--

- Webサイトの構造をツリー方式で簡略化して表示
- サイトを限定したカテゴリ検索と考えることができる
- サイト内検索と合わせて利用するとよい
- 例：国立情報学研究所公式Webサイトのサイトマップ
 - http://www.nii.ac.jp/top/sitemap_jp

特定のメディアに限定した検索

- 特定のメディアに限定した検索サービスがいろいろある。画像、ブログ、ニュースなど。
- 例：Google画像検索
(<https://www.google.co.jp/> の「画像」をたどる)
 - 画像に限定した検索ができる
 - p.51【例題7】Webを通して提供されている「空」の画像にはどんなものがあるか。
 - Google画像検索を使って「空」を検索
→ **画面イメージ**をpptへ

演習：好きなテーマで 2つの検索エンジンを検索して比較

検索エンジンによって検索結果が大きく変わることを実際に検索してみて実感してみよう！

- 自分の好きなテーマで、**2語の検索語**（例：「日本 滝」）を使って検索（検索デスクを使う）、**「日本」、「滝」以外の検索語を使う**こと
- **bing**と**Google**を両方を同じ検索語で検索
 - Yahoo!とGoogleは現在おなじ検索エンジンをつかっているため、Yahoo!のかわりにbingを使う
- **20件以上ヒット**する検索語を選ぶ
- 2つの検索結果を見比べてみる

第6回演習課題(1)

- bing およびGoogle において先に行った演習(2語で検索)の結果を、それぞれの上位20件ずつのWebページURLとタイトルと適合等を報告する
- 提出 ✕ 切: 次回演習開始時
- 「情報検索演習」、「第__回演習課題: 検索エンジンの比較」、「学籍番号」、「氏名」、「提出日」を忘れずに記入すること
- 4ページ以内に抑え、4分割印刷すること
 - つまり、提出するのはA4用紙1枚分!

第6回演習課題(2)

--まとめ方--

- 検索式(キーワード2語からなるもの)
- 検索実行(Webページ参照)日付
- ヒット件数:bing, Google
- 適合判定の基準(どんなものが正解か?)
- 考察(複数の検索エンジンを使う必要性)
- bingの検索結果
 - 上位20件: 順位, タイトル, URL, 適合判定(Oなど)
- Googleの検索結果
 - 上位20件: 順位, タイトル, URL , 適合判定(Oなど)
- 共通URL(あった場合): bing〇位– Google〇位の対応を記載
 - b1-G2 http://.....
 - b2-G3 http://..... など

第6回演習課題補足

- 「検索デスク」から検索する
 - <http://www.searchdesk.com/>
- bing/Googleの検索結果を保存する
 - IEで「ファイル」「名前をつけて保存」、リムーバルディスクを選択、「保存」
 - 20件なので次のページもやること
- 保存したファイルを使って、タイトル、URLのコピー、適合判定(○、×、△)をやる
 - 後日検索すると結果が変わることがあるため
- 検索実行(Webページ参照)日付は、「↑のファイルを保存した日」とする