

基礎情報処理(F)

第4回資料

Webの使い方
— サーチエンジンの利用を中心に —

担当者: 高久雅生

2007年5月12日(土)

masao@nii.ac.jp

本日のお品書き

- 前は、メールの使い方を演習した。
- 今回は、インターネット上での情報源であるWebの初歩的な利用法を確認し、演習課題を行ってもらう。
 - World Wide Web
 - サーチエンジン
- 最後に簡単なレポート課題を出す

Web

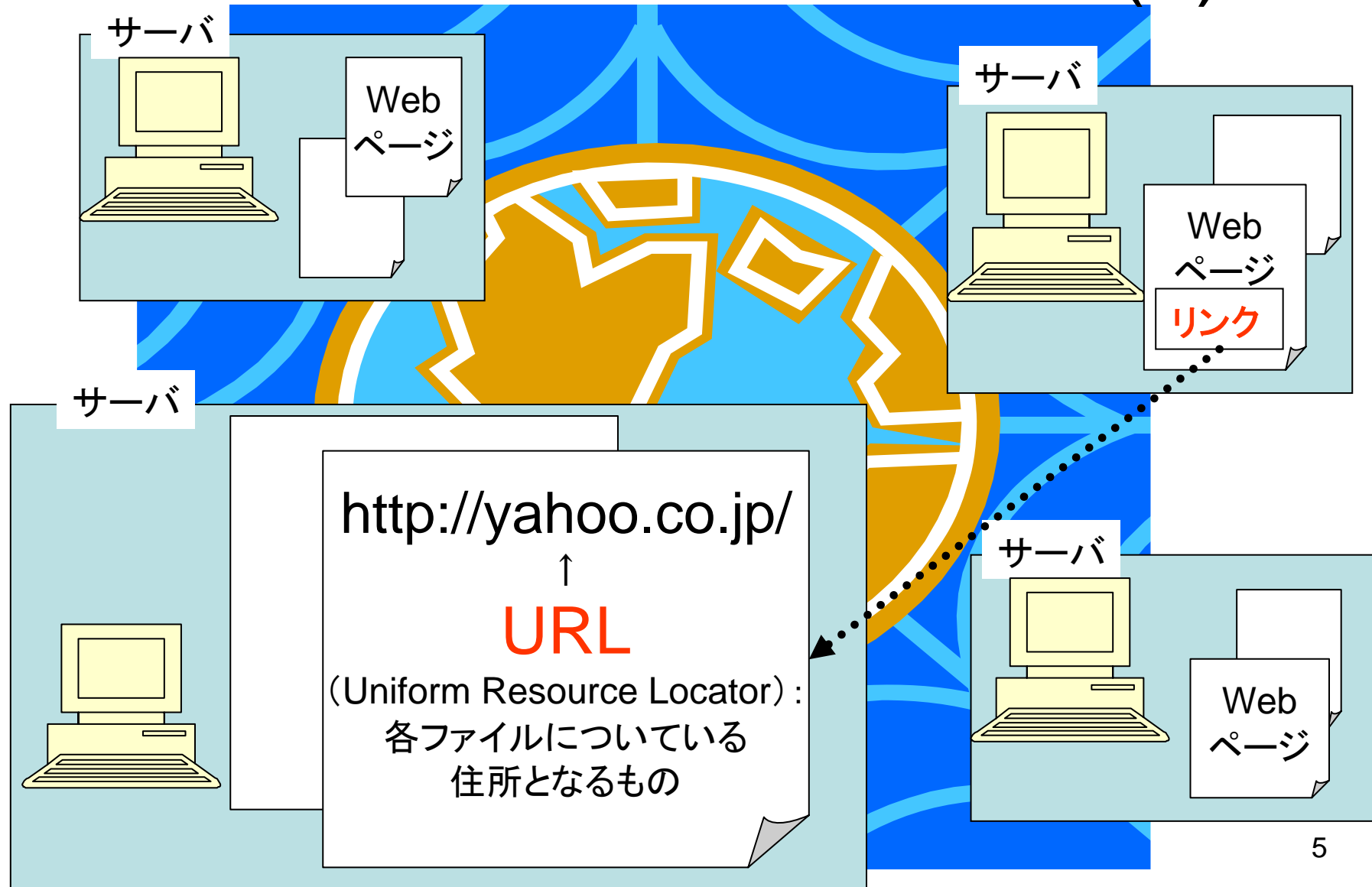
(World Wide Web; WWW)

- 世界最大の情報共有システム
- ティム・バーナーズ・リーが考案、開発
 - CERN(ヨーロッパの核物理学研究機構)に在籍していたソフトウェア科学者
- 1990年代初頭に開発され、1990年代を通じて爆発的に普及

基本的な注意事項

- いわゆる「文書」「文献」といった静的な情報だけでなく、ユーザ認証による様々なサービスが提供されています
 - ショッピング: [楽天](#), Amazon.co.jp
 - Webメール: [Yahoo!メール](#), [Gmail](#)
 - 掲示板・コミュニティ: [Mixi](#), [2ちゃんねる](#)
- これらサービスと、各種DB、サーチエンジン等を駆使すれば、多くの有用な資料が入手可能
 - 時間、場所を問わず、情報入手可能 → 電子図書館
- ただし...
- 情報の信頼性に注意
 - リンク切れ
 - 違法情報
 - プライバシー
- Web上では入手不能な情報も数多い
 - これらは有料会員制のページだけで利用できたり、紙媒体のものでのみ利用できたりすることも多い。
 - 基本は図書館における調査です！
 - (今日はWeb上での情報検索を例に演習しますが...)

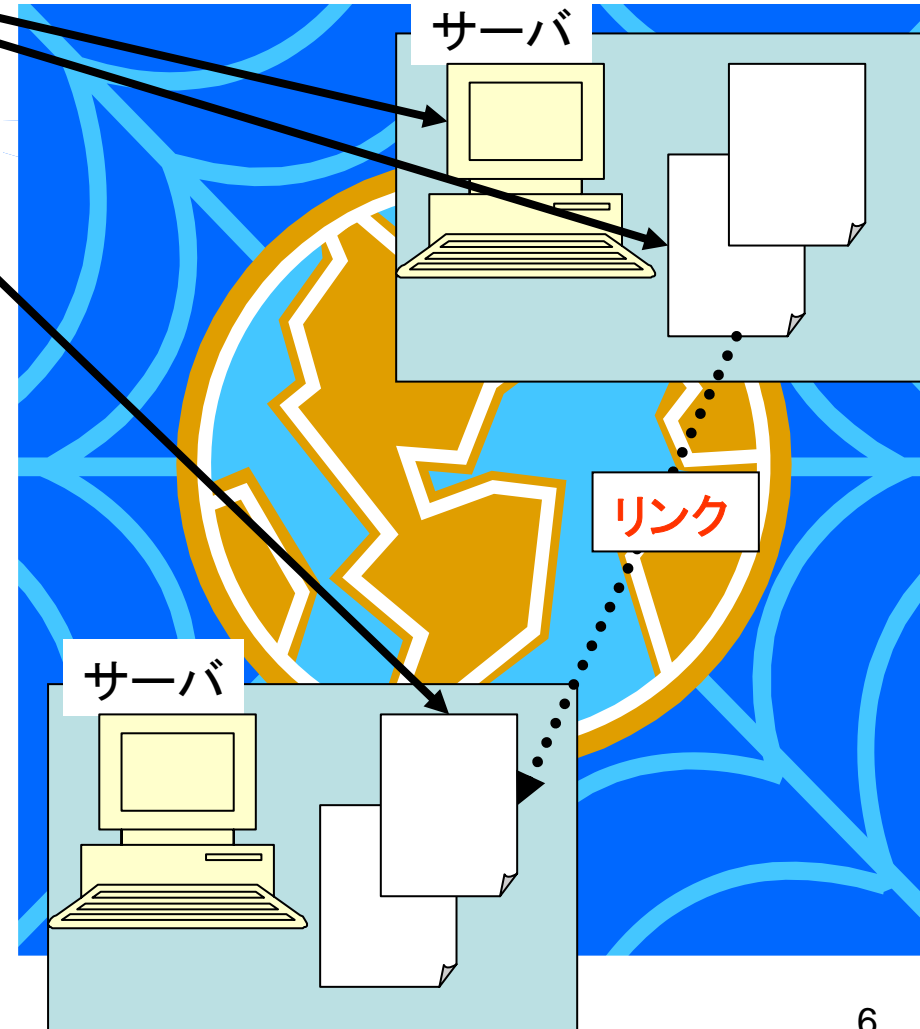
WWW: World Wide Web(1)



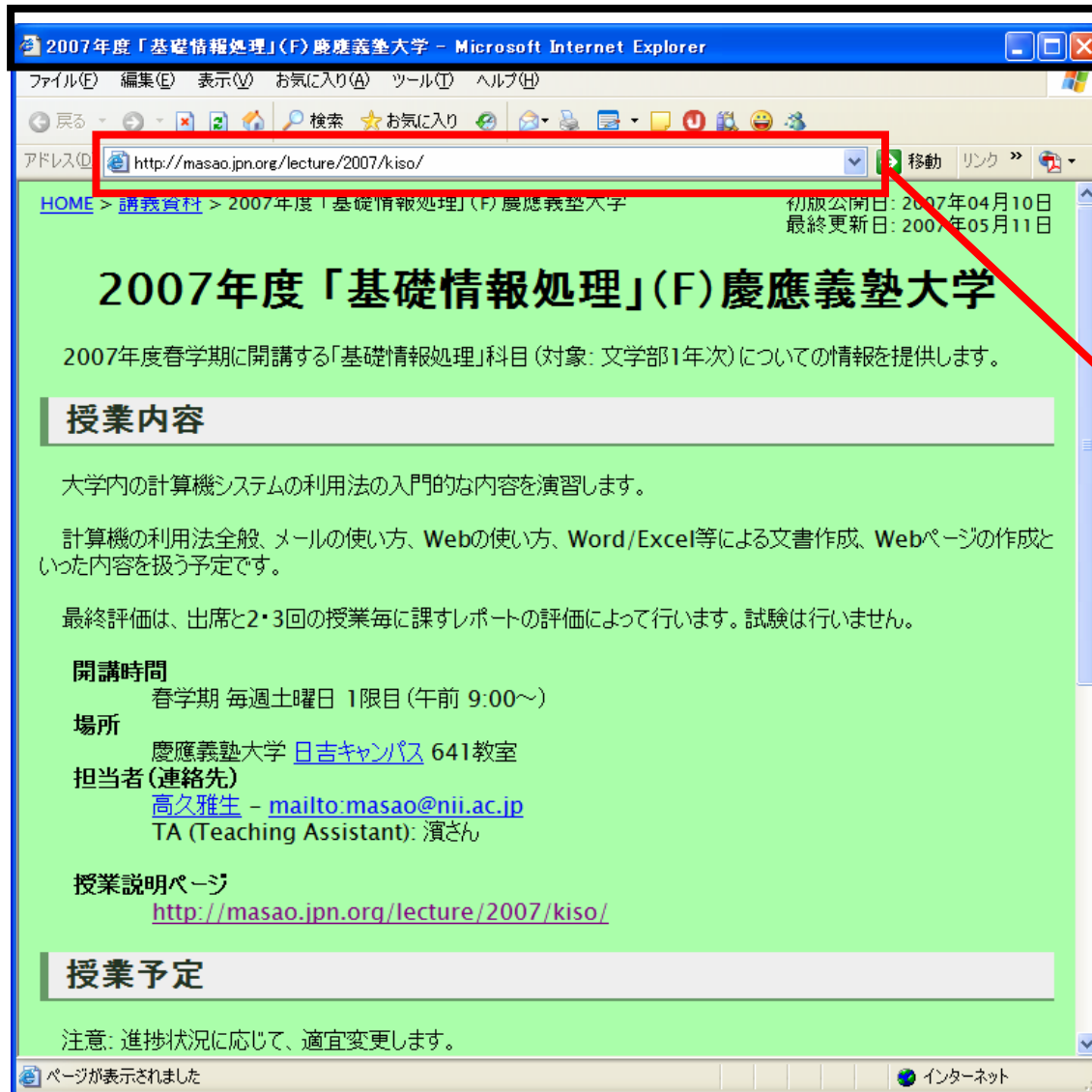


Web (2)

WWWクライアント:
WWWブラウザ:Internet Explorer



Webページ



Webページのタイトル

“2007年度「基礎情報処理」
(F)慶應義塾大学”

WebページのURL

[http://masao.jpn.org/
lecture/2007/kiso/](http://masao.jpn.org/lecture/2007/kiso/)

Webページ(2)

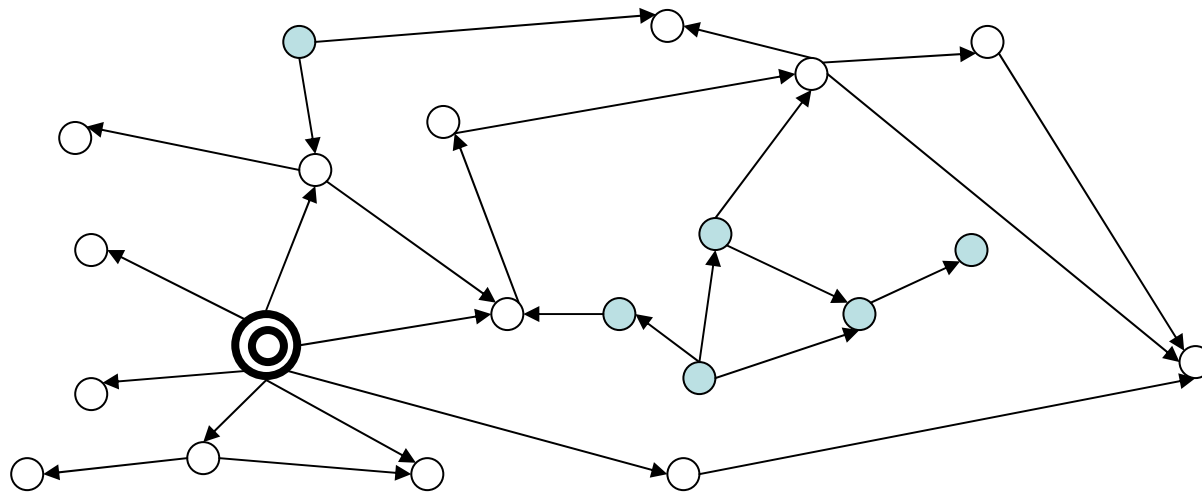


Webページのタイトル
慶應義塾ホームページ

WebページのURL

`http://www.keio.ac.jp/`

Webの模式図



出典: 三訂情報検索演習, p.38

検索エンジン

- サーチエンジンとも言う
- 種類
 - ディレクトリ型(カテゴリ型)
 - ロボット型
 - ハイブリット型
 - メタ検索エンジン
- ポータルサイトの核となる一機能

ロボット型検索エンジン

- 代表的なもの: Google, Goo
- ロボット(収集プログラム)でページを収集
 - 登録しているサイトは膨大
 - 検索結果の精度が低くなりがち
- ランキング(スコア)
 - 関連の高いものから順に表示する

ディレクトリ型検索エンジン

- カテゴリ型検索エンジンともいう
- 代表的なもの: Yahoo! カテゴリ
- 人間がページを収集
 - 一定の収集基準がある
 - 有用なサイトが得られる確率が高い
 - ロボット型に対して登録されているサイトが少ない
- 人手でページの説明情報を作成
- 人手で適切な階層(カテゴリ)に位置づけ
 - 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
- 検索システムは階層や説明情報も検索

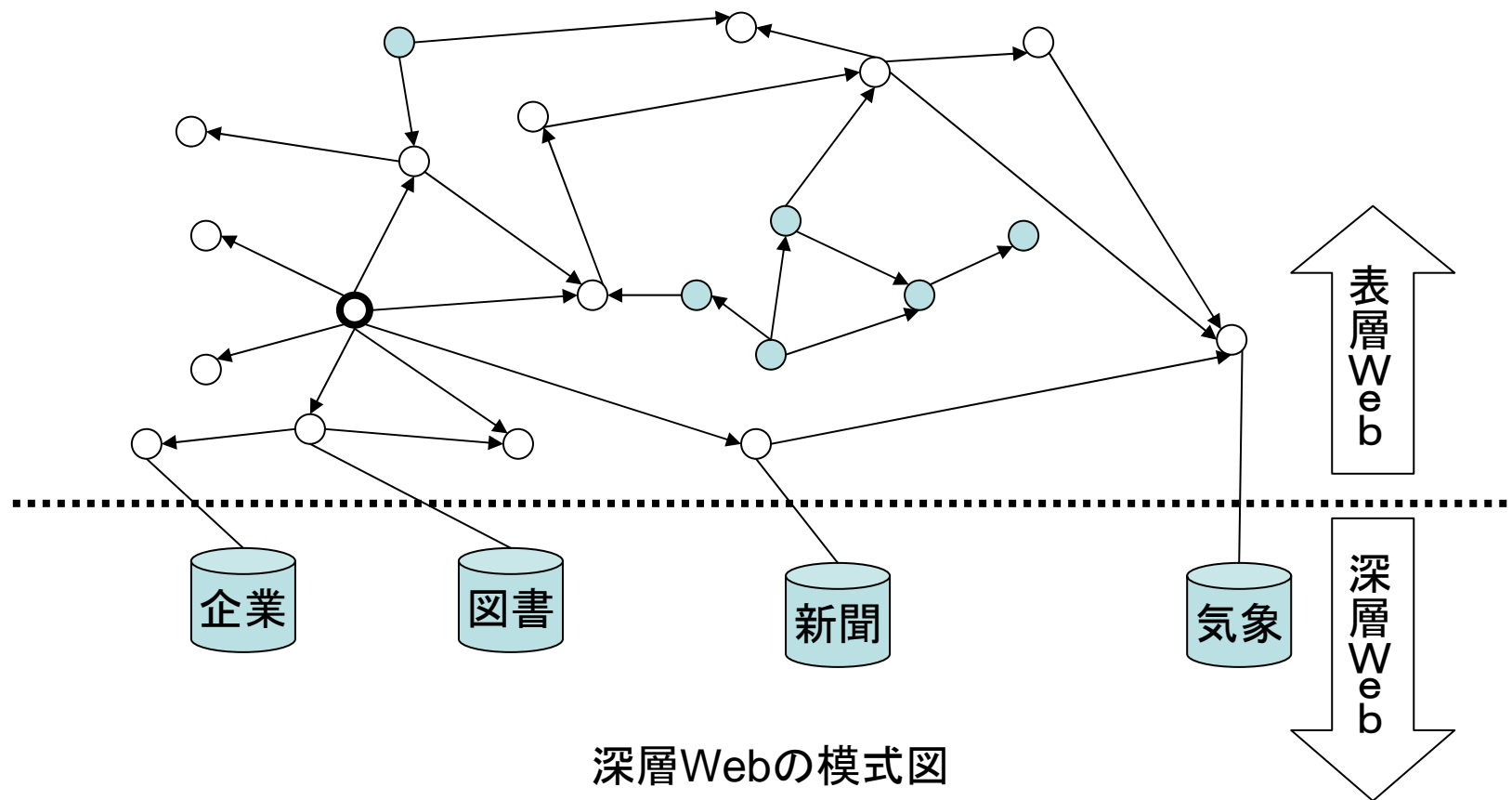
ハイブリット型検索エンジン

- ディレクトリ型とロボット型を組み合わせたもの
 - 最近はほとんどがこのタイプ
 - 大抵どちらかのタイプが優先的に
 - Yahoo! はディレクトリ型優先のハイブリット型
 - ロボットで収集したサイトも検索できる。
 - Googleはロボット型優先のハイブリット型
 - 「ディレクトリ」で、カテゴリでブラウジングできる。
 - 「ディレクトリ」のリンクをたどってみよう

検索エンジン利用の一般的な注意

1. 表層Webしか検索できない
2. 検索結果の質や信頼性を確認する
3. ヘルプを活用する
4. 検索エンジンによって結果が大きく異なる
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、記号は半角を使う

表層Webと深層Web



深層Webの模式図

出典: 三訂情報検索演習, p.59

Yahoo! Japan:

<http://yahoo.co.jp/>

- 日本でよく利用されている
- 人手による分類を行っていると言われる
- 日本人に有用と考えられるページを収集
- 簡単な解説をサイト毎に付与
- Yahoo!カテゴリ
 - 検索例1: カテゴリを辿って「国立教育政策研究所の教育図書館」を探す
- 登録サイト検索
 - 検索例2: 「ユネスコ」を検索
- Web検索
 - 検索例3: キーワード検索: 「島津製作所 ノーベル賞」
 - 検索オプション(詳細検索)

Google(1):

<http://www.google.co.jp/>

- Google: 検索式
 - AND検索: 「半角スペース」
 - OR検索: 「OR」
 - Not検索: 「-」
 - フレーズ検索: 「”」でくくる
- 「I'm Feeling Lucky」 ボタン
 - 最高位の検索結果を表示
 - 検索例4: 「気象庁」

Google (2)

- 検索オプション(詳細検索)
 - 語の区切りには半角を使うこと
 - 論理演算を使った検索
 - ページの記述言語の指定
 - 情報のファイルタイプの指定: doc, 画像ファイル
 - 情報の更新日付
 - ドメイン指定
 - 検索式で書けるものもある

Google (3)

- 検索例5

- 久留米出身の洋画家で坂本繁二郎を除く青木繁、古賀春江のサイトを検索する

- 検索例6

- 図書館と著作権に関する、日本語のページで、日本の政府省庁(gov.jp)のページを検索したい。

Google (4)

- イメージ検索
 - 画像が検索できる
 - 検索例7:「空」

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も実感してみよう！

- 自分の好きなテーマで、**2語の検索語**（例：「慶應大学 日吉」や「日本 滝」）を使って検索
- **Yahoo! Japan**と**Google**を両方を同じ検索語で検索
- **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

課題1-1

- Yahoo! Japan および Google において同一の検索式で検索を行い、それぞれの上位10件ずつのWebページURLとタイトルを報告する
 - 検索式: キーワード2語からなるもの
 - 結果のまとめと考察:
 - 検索テーマ、適合判定の基準、考察など
 - Yahoo! Japanの検索結果
 - 上位10件: タイトル, URL
 - Googleの検索結果
 - 上位10件: タイトル, URL
 - 共通URL(あった場合): Yahoo!○位 – Google○位の対応を記載
 - Y1-G2
 - Y2-G3など
- Webページの参照日付をつけること(課題の遂行日)
- テキストファイル `report1-1-学籍番号.txt` としてメールに添付する

課題1-2

- 以下からの選択課題：
 - A:「2007年夏にヨーロッパで開催されるSIGIR2007という国際会議に参加したいので、会議のスケジュール、場所を知りたい。また、開催都市周辺にお勧めの観光地などもあれば知りたい。サーチエンジンなどから検索し、関連する情報を集める。」
 - B:「テーマ(浜崎あゆみ)として、生年月日、デビュー時期、ファンの年齢層、所属レコード会社を調べる。また、所属会社については、彼女の他に所属しているアーティストの代表例、会社の設立時期についても知りたい。サーチエンジンなどから検索し、関連する情報を集める。」
 - ※()内のテーマ人物等は別のアーティストに変更しても良い。変更した場合はその旨を明記すること。
 - なお、単なる回答だけでなく、どのようにしてその情報を集めたか、そのプロセスを記すこと(こちらの方が重要!)。参考文献はできるだけ信頼のおける情報であることに留意すること。信頼できない場合は複数の情報源によって確認すること。
- 入手した情報を10行程度までの簡潔な日本語にまとめる。
- 参考文献(参照したWebページ群):
 - URL, タイトル, 参照日付
- プロセス:
 - 使用したサーチエンジン、その他サービス
 - 使用した検索式(キーワード)他
- テキストファイル `report1-2-学籍番号-[AB].txt` としてメールに添付する

課題1・提出方法

- メールにて提出
 - 宛先: masao@nii.ac.jp
 - 件名: [\[kiso-f report1\] 学籍番号](#)
 - (上記の件名はすべて半角英数字にしてください)
- 締め切りは次週の授業日前日: 2007年5月18日(金)
- 本文
 - 氏名、学籍番号を明記すること
 - 課題1-1において実行した検索式を明記すること
 - (結果自体は添付ファイルに入れる)
 - 課題1-2についても、選択課題(ABの別)を明記すること
 - レポート課題についての感想
- 課題1-1:
 - 作成したテキストファイルを添付
- 課題1-2:
 - 作成したテキストファイルを添付
 - ※テキストファイル名はすべて半角英数字とすること。
 - ※作成したテキストファイル中にも、氏名、学籍番号は明記のこと。

まとめ

- Webにおける検索エンジンの利用法
 - レポート課題1（来週の授業までメールにて提出）
- 次回はWord演習を行う予定です

本日の出席

- 授業担当者(高久)宛に、メールを送ってください:
 - 宛先は masao@nii.ac.jp
 - 件名(Subject)は [\[kiso-f 20070512\] 学籍番号](#)
 - (上記の件名はすべて半角英数字にしてください)
- 内容:
 - 氏名
 - 学籍番号
 - 今日の授業の感想、要望

参照文献、参考文献

- 本日の資料には国立教育政策研究所・研究員・江草さんによる授業資料の一部を利用させて頂きました:
 - － 江草由佳: “平成17年度後期: 情報検索演習”.
<http://momiji.mimoza.jp/lecture/2006/ir/>
(参照 2007-05-10)
 - － 原田智子編著ほか. 三訂情報検索演習. 樹村房. 2006.
203p. (ISBN 4-88367-131-3) 第2刷
- 読み物: World Wide Web開発者による
 - － ティム・バーナーズ・リー. 高橋徹 (訳): “Webの創成 — World Wide Webはいかにして生まれどこに向かうのか”.
毎日コミュニケーションズ, 2001, 279p. (ISBN 4839902879)