

情報検索演習

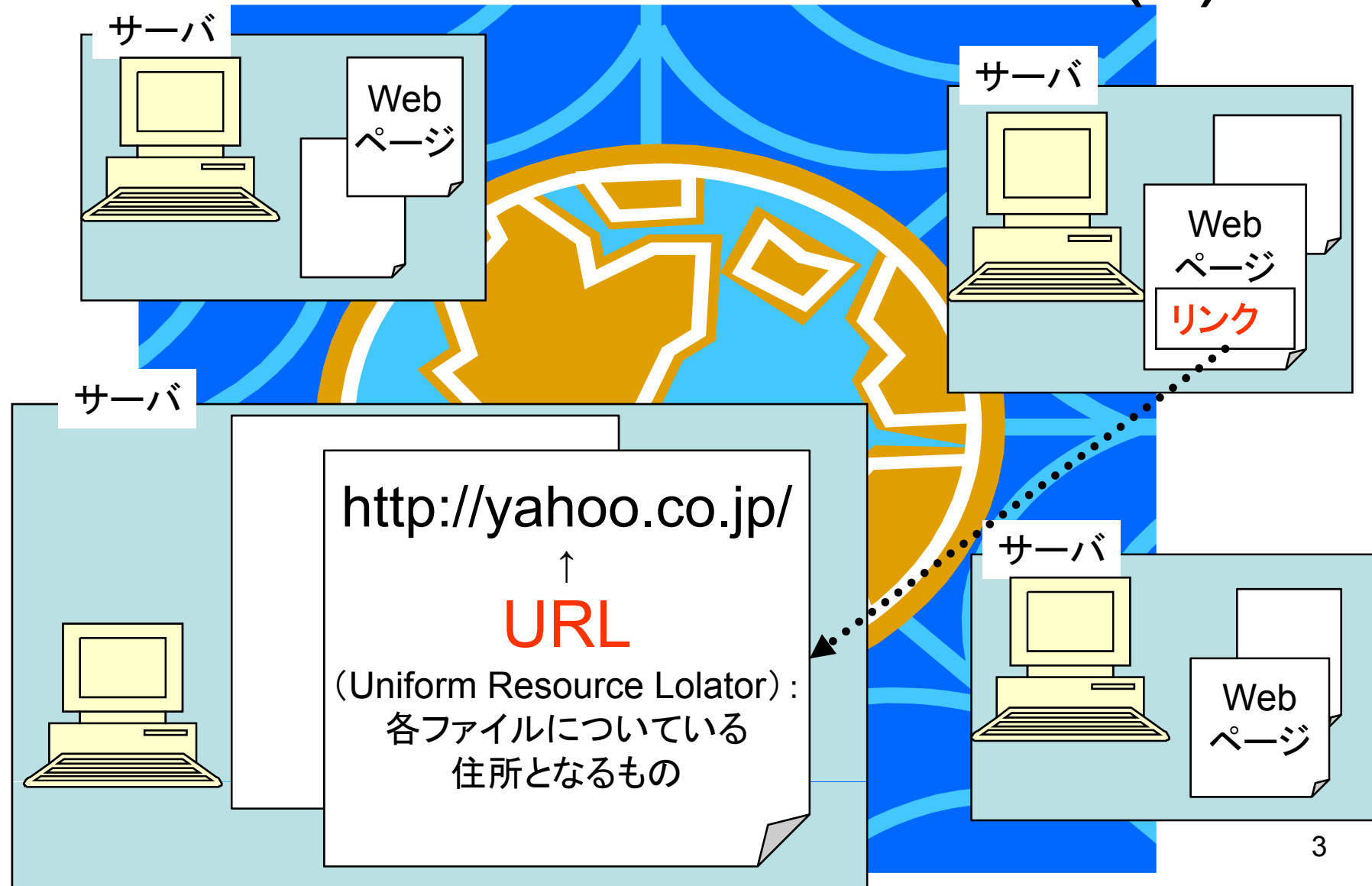
第5回資料

鶴見大学
江草由佳
国立教育政策研究所
yuka@nier.go.jp

本日のお品書き

- 第1回レポート課題提出
- 演習: Webページ、Webサイトの探し方
 - 検索エンジン (Yahoo! Japan、Google)
 - メタ検索エンジン (検索デスク、Ceek.jp)
- 第6回演習課題

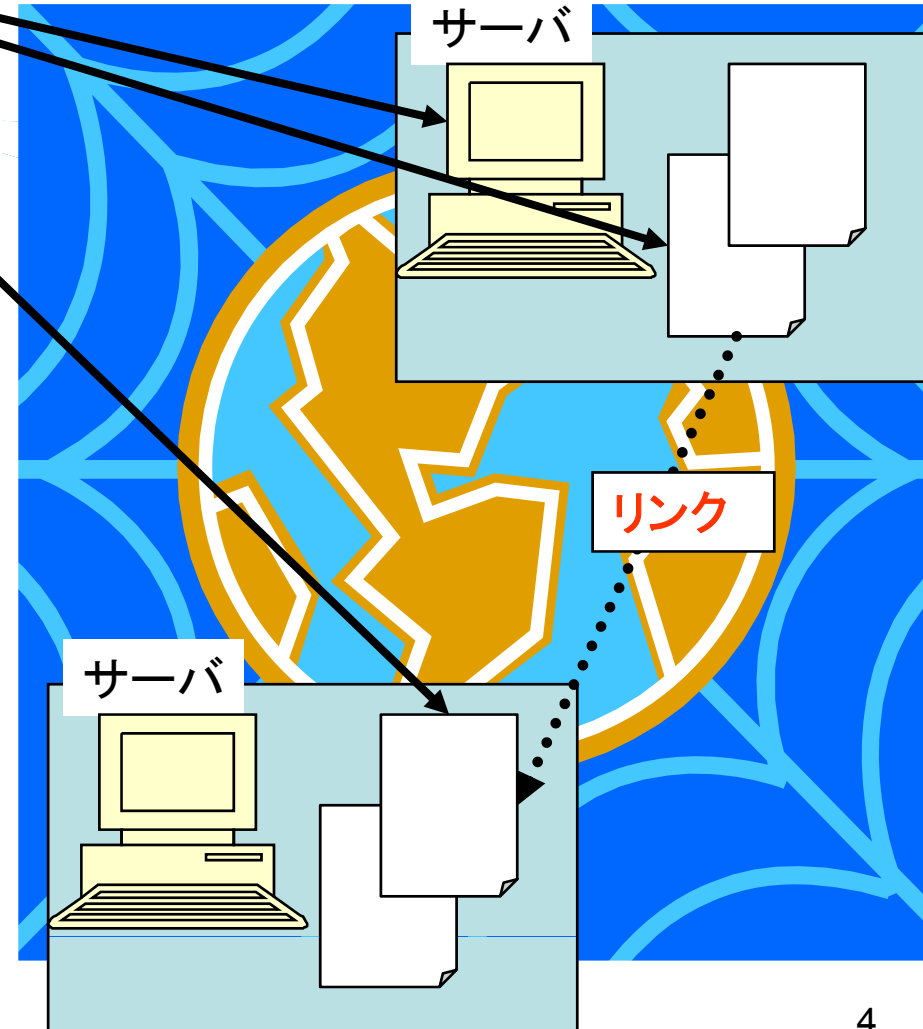
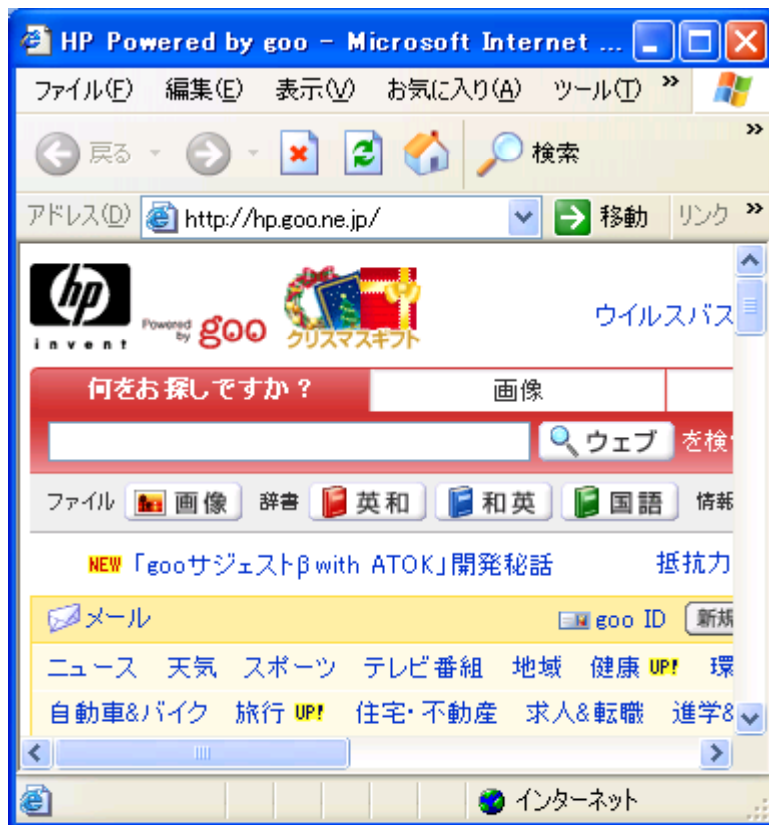
WWW: World Wide Web(1)





Web (2)

WWWクライアント:
WWWブラウザ: Internet Explorer



Web(3)

- Webページ
- Webサイト
- ホームページ、トップページ
- 検索エンジン
- アーカイブサイト
- サブジェクトゲートウェイ、リンク集
- サイト内検索、サイトマップ
- ポータルサイト

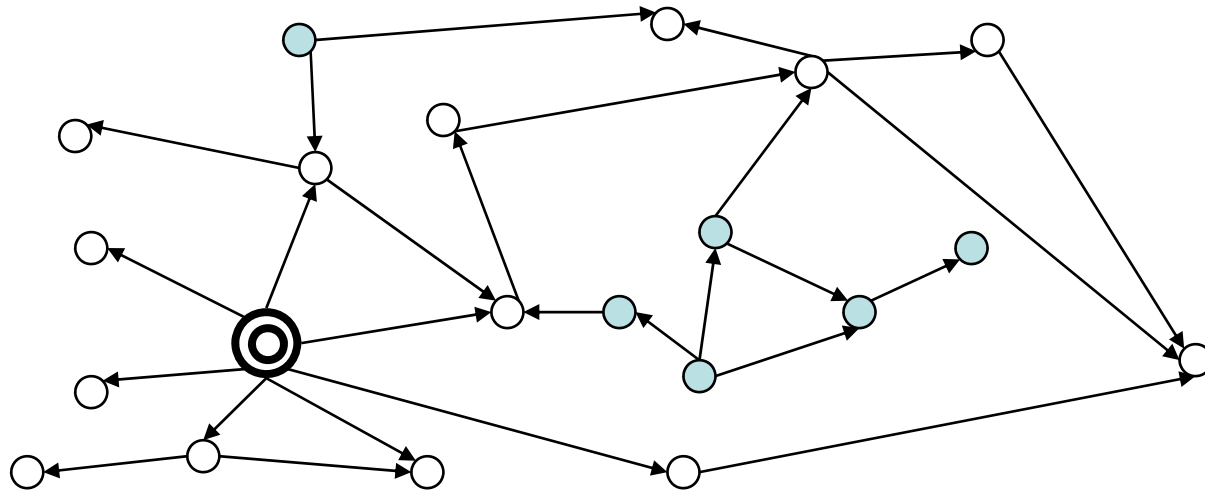
検索エンジン

- 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

p.38 3-1 図 Webの模式図



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

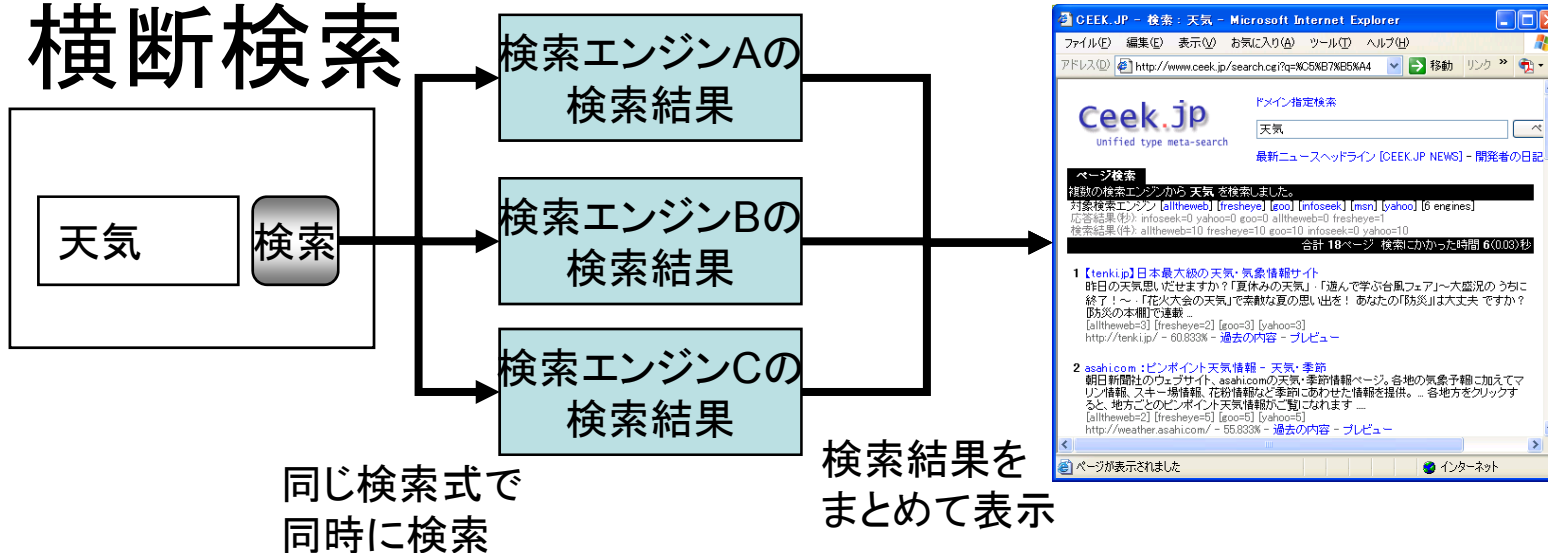
ハイブリット型検索エンジン

- ディレクトリ型とロボット型を組み合わせたもの
 - 大抵どちらかのタイプが優先的に

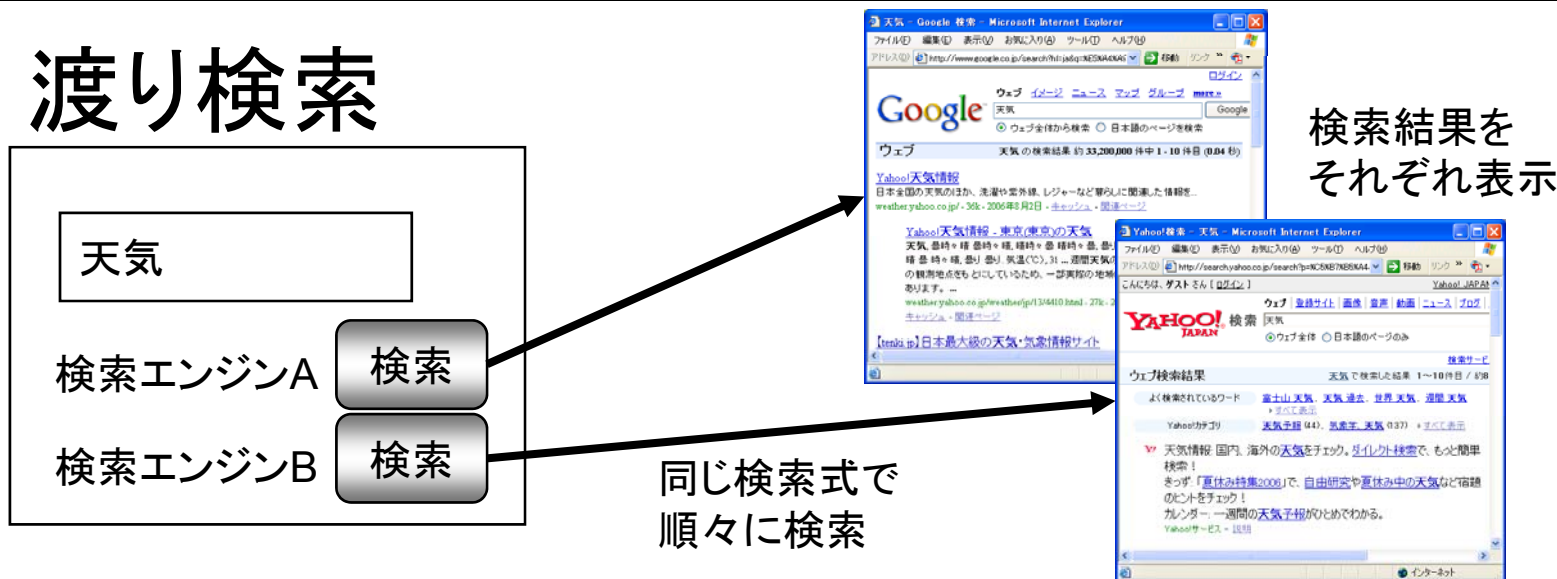
メタ検索エンジン

p. 40 3-2図 メタ検索エンジンの種類

横断検索



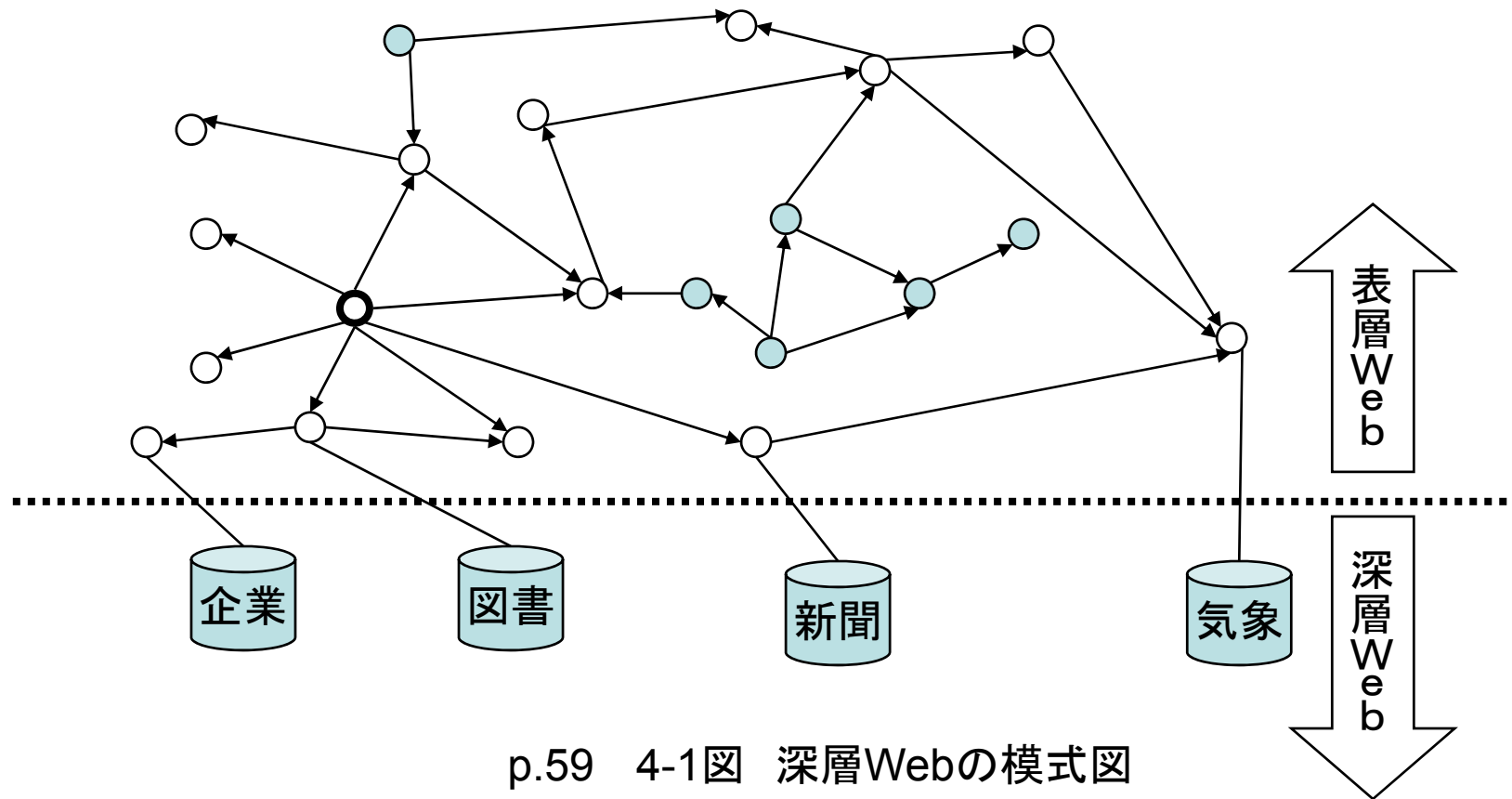
渡り検索



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

表層Webと深層Web



p.59 4-1図 深層Webの模式図

Yahoo! Japan:

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

- 日本でよく利用されている
- アナリストは200人といわれている
- 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
 - リンクをたどって「教育図書館」← p.44 検索例-1
- 登録サイト検索
 - 「ユネスコ」を検索 ← p.45 検索例-2
- Web検索
 - キーワード検索:
「島津製作所 ノーベル賞」 ← p.45 検索例-3
 - 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
 - 演習 : 「気象庁」 p.47 検索例-4

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)

- p.48 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
- P.49 検索例-6
 - 図書館と著作権に関する、日本語のページで、日本の省庁(go.jp)のページを検索したい。

Google (4)

- 画像検索
 - 画像が検索できる
 - 演習:「空」 p.49 検索例-7

メタ検索エンジン

- 複数の検索エンジンを同時に/並行的に検索
- 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
- 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

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：「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