



## 情報検索演習

### - 検索エンジン -

2015  
時実 象一  
tokizane@aichi-u.ac.jp

1

## Yahoo! と Google



- Yahoo!
  - スタンフォード大学の Jerry Yang と David Filo が開発 (1994)
  - Web ページを体系的に整理して提供
  - ディレクトリ (カテゴリ)
- Google
  - スタンフォード大学の Larry Page と Sergey Brin が開発 (1998)
  - 高速の検索エンジンに特徴

2



## 検索エンジンの種類

- ロボット型
  - ロボットがインターネット上の Web ページを調べて索引
  - 見つかったページ上のすべての文字情報が検索できる
  - Google, Yahoo!, MSN などほとんどの検索エンジン
- ディレクトリ型
  - 人手で Web ページを整理・登録
  - Yahoo! カテゴリ

3

## ロボットの対象



- Web ページ
  - ページに文字で書かれている情報
- 対象でないもの
  - Deep Web
    - データベース (OPAC など) の中身
    - ダイナミックに作成されるページ
    - ロボットが禁止されているページ
      - <meta name="robots" content="noindex,nofollow">

4



## Google の検索結果

Google 検索結果 約 6,790,000 件 (0.05 秒)

スニーカー 市場 ショップ ナイキ アディダス ブラックフライデー

スニーカー 好き 必見!手に入らなかっアモノまで楽天で

激安価格

セ・スニーカーお探しなら

いろいろショップ・メーカーのものが 比較購入できMSN

ショッピング

【楽天市場】スニーカー カテゴリ: ナイキ、アディダス、アバランチなどの通信

... 春夏は是元でアビール・春夏は是元でアビール 老若男女、クールなスニーカーで夏を 追う!注目アイテムだらけです!スニーカーシューブラネット ... お洒落が是元だ! じもったいない!スニーカーと相性抜群なお洒落なアイテムをご紹介! ... agent rakuten.co.jp/sneaker/- 164k - 2007年6月28日。

カッショ・関連ページ

5

## Google の特徴



- 検索結果はランクの高いものから表示される
- 複数のキーワードの位置関係 (近接)
- PageRank
  - 他のサイトからリンクを数える
  - 重要なサイトからのリンクは点数が高い
  - 点数の高いサイトからのリンクは点数が高い



6



## Web ページの引用



- 匿名の情報は出典としては扱えない
  - 情報の真偽を確認できない
    - Traceability
- 引用する際は日付を明記
  - Web ページはすぐ修正されたり削除されたりするため