

情報検索演習 - 検索エンジン -

2015
時実 象一
tokizane@aichi-u.ac.jp



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

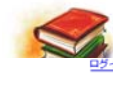
- Web ページ
 - ページに文字で書かれている情報
- 対象でないもの
 - Deep Web
 - データベース (OPAC など) の中身
 - ダイナミックに作成されるページ
 - ロボットが禁止されているページ
 - `<meta name="robots" content="noindex,nofollow">`



4

Google の検索結果

Google 検索結果のスクリーンショット。検索キーワードは「スニーカー」。検索結果には「スポンサー」のリンクと「ウェブ」の検索結果が表示されている。赤い丸は「スポンサー」のリンクを、緑の丸は「ウェブ」の検索結果を囲んでいる。



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

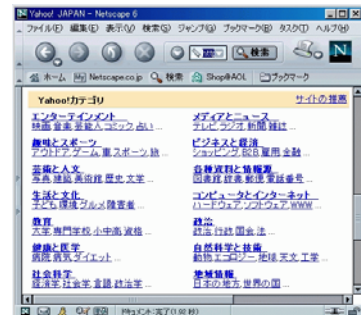
Yahoo! カテゴリ

- ディレクトリ
 - 登録サイトが分類されている



7

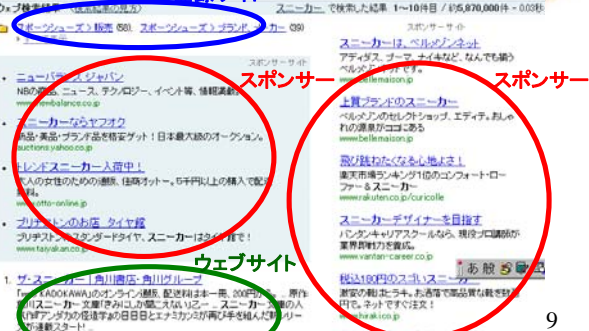
Yahoo! カテゴリ



8

Yahoo! の検索結果

登録サイト



9

キャッシュ

- キャッシュ
 - Web ページを一時的に保存したもの
 - ヒットしたキーワードを見ることができる



このページは、Google で 2007 年 6 月 28 日 09:24:12 GMT に保存された <http://www.e-shops.jp/sh/rk155/html/> のキャッシュです。
 Google のキャッシュとはウェブページを一時保存したページのスクリーンショットを指します。
 そのため、このページの最新情報がない場合があります。こちらから最新のページ(サイト)表示を参照してください。
 このキャッシュ ページには、もはや参照できなくなっている画像が含まれている可能性があります。テキストのみのキャッシュ ページをご利用になる
 このページへのリンクやお知らせの登録には次の URL をお使いください。
<http://www.google.com/search?cache?cx&sl=ja&sa=www.e-shops.jp/sh/rk155/html/>

Google はこのページの所有者と何ら関係性はなく、そのコンテンツについても責任を負いません。

ハイライトされているキーワード: スニーカー

10

Google の特徴 – AND, OR, NOT

- 検索語を並べると AND の検索となる
 - 「豊橋 大学」
- “-” を使うと NOT 検索となる
- OR は「豊橋 OR 渥美」のように検索
 - 小文字の or は駄目



11

検索エンジンのしくみ

- 何が上位に来るか
- SEO
- 食べログの仕組み
- アフィリエイト
- ネット世論調査
- なぜ見つからないか



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