

5-1 ロボット型検索エンジンの作成に関する基本研究

2年29番 與羽 祐也

指導教員 ソソラ

1. はじめに

インターネットの普及により、多くの情報をウェブページから得られるようになった結果、必要な情報を探すために用いられる“検索エンジン”的重要性がますます高まっています。そこで、検索エンジンの動作やプログラムがどのようなものなのか具体的に理解し、小規模な検索エンジンを動かしてみたいという理由で本テーマを選定しました。

本研究の目的は検索エンジンの仕組みを理解し、検索の対象となるウェブページを自由に設定できるような自分専用の My 検索エンジンをパソコン上で運行してみることです。

以下、まず検索エンジンの仕組みを説明し、前回の発表での研究計画（参照 3 章）において得た研究成果と今後の課題について述べます。

2. 検索エンジンの仕組み

私たちに馴染みの深い検索エンジン「Google」の場合、ロボットやクローラーと呼ばれる検索エンジンプログラムがインターネット上を巡回し、ウェブページの情報を収集します。

そして、その収集したウェブページの中からユーザーが指定したキーワードと関連性の高いページが順に表示されます。

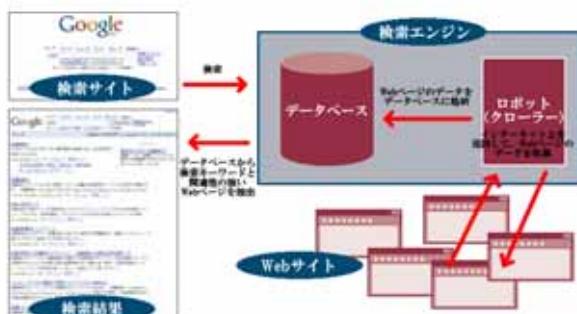


図 1 Google の大まかな仕組み

3. 研究計画

前回の中間発表での研究計画は以下の通りでした。

- 開発環境の準備
- 参考文献に基づき小規模な検索エンジンを動かし、動作の確認
- 索引用のテーブルの構造
- オブジェクト指向 Ruby 言語の基礎

3.1 開発環境の準備

開発環境の準備として以下の順番で、参考文献 2 の通りインストールを行いました。

- Apache (Web サーバー)
- MySQL
- Ruby
- Ruby + MySQL (MySQL/Ruby for Windows)
- Chasen
- Ruby + Chasen (Ruby/Chasen Alternative)

3.2 検索エンジンの動作の確認

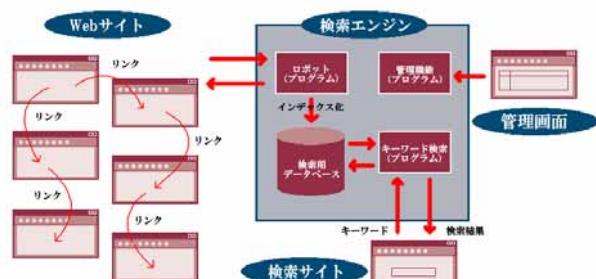


図 2 本検索エンジンの仕組み

小規模な検索エンジンでは一般的な検索エンジンと違い、お気に入りのウェブページを検索対象として設定します。

検索エンジンのプログラムが検索対象のページのみを巡回し、ウェブページの情報を収集します。

収集した HTML ファイルからタグ部分を取り除き、検索対象となる文字列を取り出します。文字列を単語に分解したものと、それを含む HTML ファ

イルの URL をデータベースに保存します。その単語が URL を求めるための索引の役目を果たします。

検索エンジンは 索引となるデータベースを検索して、指定されたキーワードに一致する単語を含む HTML ファイルの URL を特定します。

以下のステップによりエンジンの動作の確認を行いました。

ステップ 1. データベースやテーブルの作成

- ログイン ■ サイトの熱風属性 ■ 記録サイト一覧 ■ 証跡サイト一覧 ■ 検索サイト
- システム管理
 - ・ インターネットの巡回を停止
 - ・ データベースの作成
 - ・ テーブルの作成
- ・ 登録サイトのバックアップ
- ・ 登録サイトの復元
- ・ テーブルの削除
- ・ データベースの削除

ステップ2. サイトの登録

ステップ3. ロボットの稼動

ステップ4. データベースへ蓄積

Maintenance Mode	
■ログイン	ログイン回数
■ログイン失敗	失敗回数
■登録	登録回数
■システム状況	
• 登録完了ページ	<ul style="list-style-type: none">◦ Level 1 -- 5◦ Level 2 -- 15◦ Level 3 -- 12◦ Total -- 32
• 登録エラーページ	<ul style="list-style-type: none">◦ Level 3 -- 1◦ Total -- 1
• キーワード数	<ul style="list-style-type: none">◦ Total -- 19431

ステップ5. キーワードを検索



ステップ6. 検索の結果を表示

Search Mode

■ 2000 ■ 2001 ■ 2002 ■ 2003 ■ 2004 ■ 2005 ■ 2006 ■ 2007 ■ 2008 ■ 2009 ■ 2010 ■ 2011 ■ 2012 ■ 2013 ■ 2014 ■ 2015 ■ 2016

Sort by Date | Sort by Title | Advanced Search

Asian Recipes at Cook's Country [Part 3] - episode 155

This week's episode of Asian Food Report explores Asian-inspired favorites (chicken teriyaki, pad thai) that aren't necessarily Asian. It also features a few traditional Asian dishes (such as dumplings) made with non-traditional ingredients (like bacon). The episode ends with a special segment on how to make the perfect pad Thai.

1. Korean Recipe: Makguksu, the Korean Noodle Salad - episode 156

Cooking Light's Diane Lake (Dishin') talks with Diane Englehart about the Korean classic (with 1/2 dozen gochujang) with 1/2 dozen gochujang (1/2 cup kimchi) in its present day. (Previously, it was known as "Lugukse" with 1/2 cup radish and seaweed.) (With 1/2 cup kimchi and 1/2 cup radish.) (With 1/2 cup radish.) (With 1/2 cup radish.) (With 1/2 cup radish.) (With 1/2 cup radish.)

2. Japanese Laundry - episode 157

Engagement Ring Light is here for a post-wedding interview. Diane Geng, A-Z Foods' Lisa French, Shopping Queen and Diane's Gymnastics Biggest Fan (and Laundry Queen) are here to help. Diane's Laundry Room is a hit! Diane's Laundry Room is a hit! Diane's Laundry Room is a hit!

3. Perfect Paella - episode 158

Engagement Ring Light is here for a post-wedding interview. Diane Geng, A-Z Foods' Lisa French, Shopping Queen and Diane's Gymnastics Biggest Fan (and Laundry Queen) are here to help. Diane's Laundry Room is a hit! Diane's Laundry Room is a hit!

4. Perfect Paella - episode 158

Engagement Ring Light is here for a post-wedding interview. Diane Geng, A-Z Foods' Lisa French, Shopping Queen and Diane's Gymnastics Biggest Fan (and Laundry Queen) are here to help. Diane's Laundry Room is a hit! Diane's Laundry Room is a hit!

4. おわりに

前回の中間発表での研究計画通りに開発環境の準備や参考文献に基づき小規模な検索エンジンを開発しました。

今まで文面でしか理解していなかった検索エンジンの仕組みをプログラムに触れることによってより理解が深められたと思います。しかし、卒業研究の時間を授業の課題やその他の用事に使うことが多かったために計画通りに進められなかつた部分もありました。

そこで、今後はまだ理解が浅い索引用のテーブルの構造やオブジェクト指向 Ruby 言語の基礎に集中して取り組みたいと考えています。

5. 参考文献

1. “体系的に学ぶ検索エンジンのしくみ”, 2008
 2. “Rubyで作る検索エンジン”, 2009
 3. “Web検索エンジンGoogleの謎”, 2004
 4. “Windows自宅サーバーの作り方・活かし方”, 2002
 5. “正規表現ハンディリファレンス”, 2001