

J-21 ロボット型検索エンジンの作成に関する基本研究

29 番 與羽 祐也

担当教員 ソソラバラム バドゥジャルガル

1. テーマ選定理由

インターネットの普及により、多くの情報をホームページから得られるようになり、必要な情報が掲載されたホームページを探すため、検索エンジンの重要性がますます高まってきています。そのため検索エンジンがどのようなプログラムで作られ、どのように動作している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。

2. 研究概要

本報告書では、これまでの検索エンジンに関する調べたものを簡単にまとめました。本研究の目的は検索エンジンの仕組みを理解し、お気に入りのホームページや検索の対象となるホーム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実際に運行して試みることです。

現在、インターネット上では、さまざまな検索エンジンサービスが存在し、競合しています。検索エンジンには大別してロボット型とディレクトリ型の2つの種類があります。

ロボット型の検索エンジンで最も有名なのは「Google」です。Google はウェブロボットと呼ばれるプログラムを使ってホームページを巡回してインデックス(索引)を作成し、そのデータを元に検索を行います。ただしロボット型エンジンは、全ての作業が機械的に行われているため、利用者にとって有用な情報を含んだホームページとそうではないページが一緒に検索結果に出てしまいがちです。

それに対してディレクトリ検索エンジンでは登録されているホームページは人間による審査を受け、登録にふさわしいと見なされたものだけです。ディレクトリ型の検索エンジンで最も有名なのは「Yahoo」です。ただし、人の手によるチェックを行うため、登録されているページ数はロボット型の検索エンジンと比べるとはるかに少なく、また登録作業をしていないホームページは検索の対象にならないという問題があります。

しかし、最近では2つの異なる検索エンジンの仕組みを組み合わせることで質の高い検索機能を提供しようとする混在型が多数を占めています。

このため、特定のテーマに対して、定期的に更新されているかつ有利な情報が掲載されたホームページのリンクを持っているページを検索対象として設定し、ロボット型の検索エンジンのように自動的に対象ホームページにアクセスし、データを回収することにより少しでも質の高い検索結果が得られるのではないかと考えています。

参考文献

1. 神崎 洋治 “体系的に学ぶ検索エンジンのしくみ”, 2008
2. 星澤 隆 “Ruby で作る検索エンジン”, 2009
3. 水野 貴明 “Web 検索エンジン Google の謎”, 2004
4. 武井 一巳 “Windows 自宅サーバの作り方・活かし方”, 2002

表 1 開発環境

言語	Ruby
OS	Windows XP
データベース	MySQL
Webサーバ	Apache
形態素解析SW	Chasen

3. 研究計画

表 2 9月までの研究計画

時期		研究計画
8月	下旬	<ul style="list-style-type: none"> ● 開発環境の準備 ● ウェブサーバのインストールや設定方法
	上旬	<ul style="list-style-type: none"> ● 参考文献 2 に基づき小規模な検索エンジンを動かし, 動作の確認を行う ● 索引用のテーブルの構造を理解する
	下旬	<ul style="list-style-type: none"> ● オブジェクト指向 Ruby 言語の基礎を学ぶ

4. その他

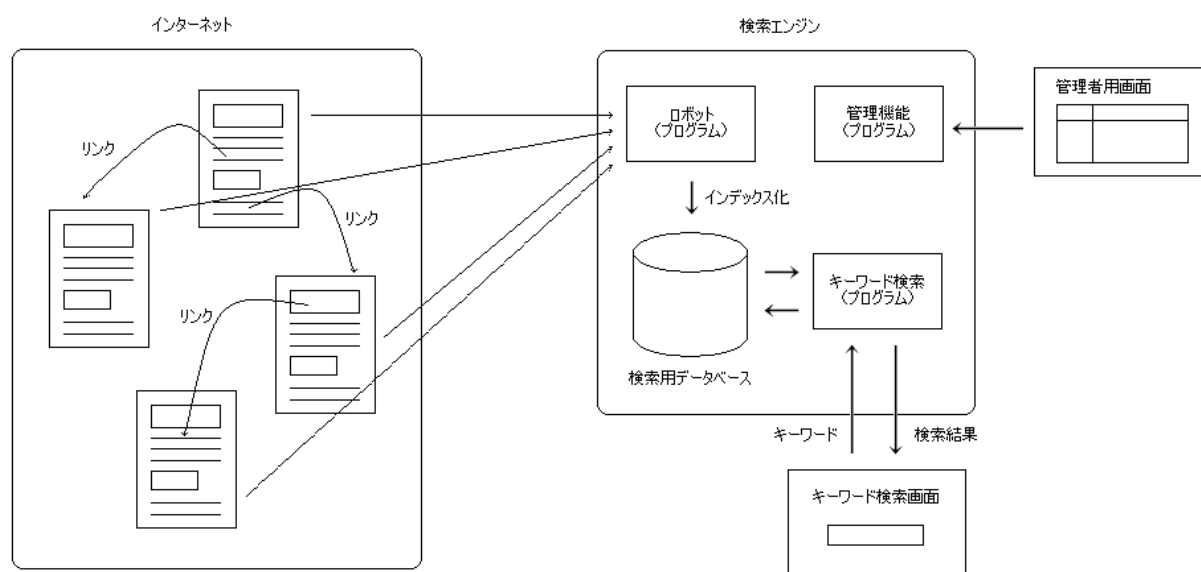


図 1 ロボット型検索エンジンの仕組み