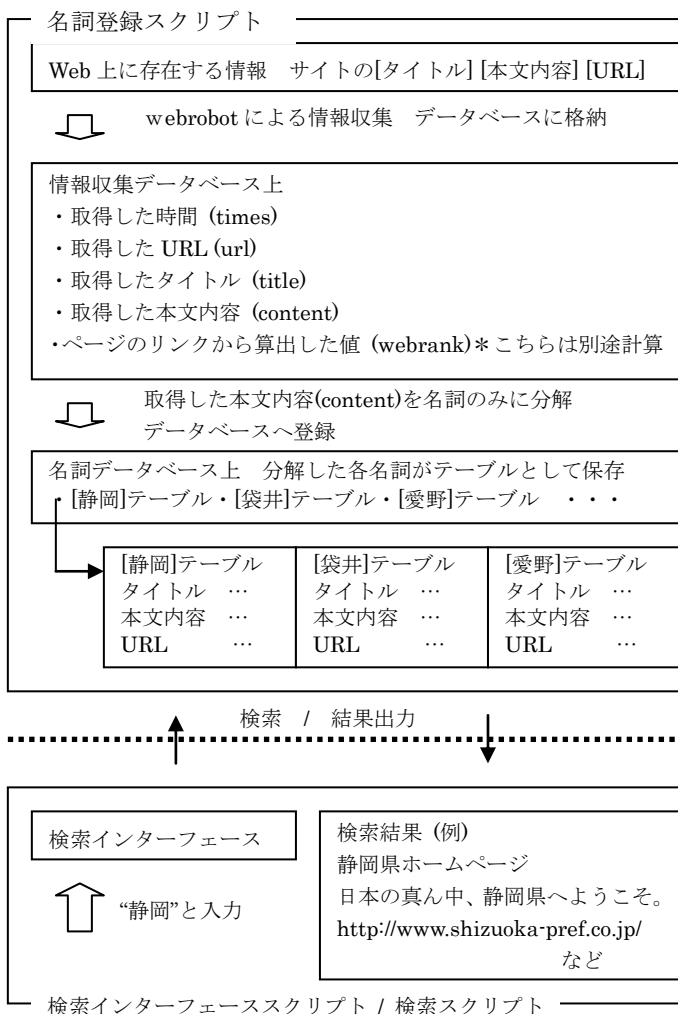


1 研究目的

今年度はサーチエンジンの検索速度向上ということに主眼を置き研究を進めた。昨年度まではサーチエンジンの骨格部分を製作するということが研究テーマが置かれておりユーザーの視点に立った際、検索結果の表示速度についてはあまり考えられていなかった。OR 検索、AND 検索などのオプションは検索サイトを使うユーザーには必須の機能であり、重要ではあるが、その後の検索速度については必要最低限の能力しかなく検索結果を表示するのにかなりの時間を要していた。それらの検索オプションを引き継ぎ、さらにユーザーが使いやすい検索エンジンに強化した。

2 システムの概要

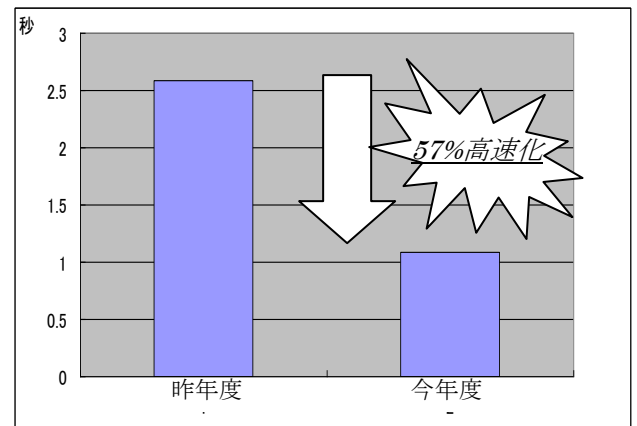
今回作成したシステムの全体図を以下に記す。



Web上から収集したWebサイトのタイトル、本文内容、URLを一度データ情報収集用データベースへ格納する。収集したデータから本文内容を取り出し、それを名詞のみに分解する(例 静岡県へようこそという内容ならば”静岡”、”県”、”へ”、”ようこそ”のように分解される)。分解したものを名詞用データベースに各名詞テーブルとして保存する。

検索部分はWebブラウザ上の検索インターフェースで検索するキーワードを入力する。その後、入力されたキーワードから名詞のみを取り出す。等しい名詞テーブルを名詞用データベースから検索する。この時にOR検索やAND検索などの検索オプションにより処理が分岐し、目的のWebサイトの一覧が表示される。なお各スクリプトはPHPで作成し、データベースはMySQLを使用している。

3 実行結果



上図は縦軸に検索時間(秒)をとったグラフである。ヒット数が同じになるような検索キーワードを探し出し、そのキーワードで検索をしている。検索に使用したURL数は4000件、ヒット数は830件、名詞登録データベースに登録された名詞数は51744語であった。

結果は昨年度のシステムが2.53秒かかったのに対し、今年度は1.09秒と約57%検索時間が高速化された。実行した際のマシン環境、検索条件はすべて同一である。この結果は設備投資には一切費用はかかっておらず、純粋に今年度の研究による成果であると考えられる。