静岡理工科大学　理工学部　情報システム学科　卒業論文
「小規模サーチエンジンシステムのデータベース可視化ツールの制作」
0713079　平林　俊哉　[担当教員　幸谷　智紀]
2011 年　2 月 17 日(木)

1. 目的

本研究では、数年前から研究を続けている小規模サーチエンジンシステムのデータベース可視化ツールの制作を目指す。昨年までの卒業研究でキャッシュの実装、名詞テーブルの並列化と分割により、即存のサーチエンジンと同じ機能を持ったシステムを作成した。

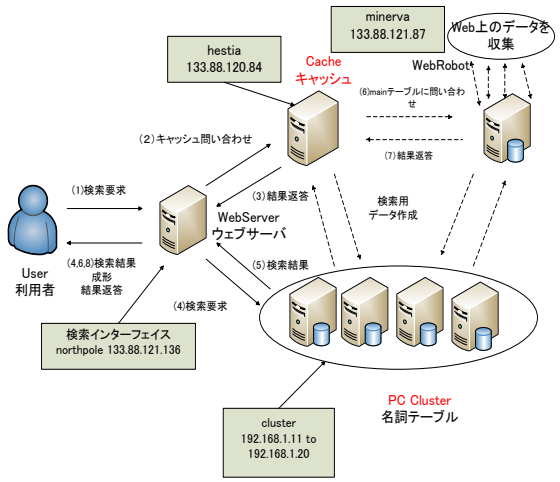
しかし、小規模サーチエンジンシステムを一般の人に紹介した時には、複雑になりすぎてシステムを理解してもらえずにいた。

そこで、今年度は、サーチエンジンシステムの状態がどのようになっているかを確認できるスクリプトを作成した。

2. システム概要

以下の図は、本研究でデータベース可視化ツールを導入した小規模サーチエンジンシステムの概要と検索の流れである。

検索に利用するデータは、WebRobot が随時収集している。PC Cluster が名詞テーブルを分散配置している。名詞テーブル作成に並行して、検索結果そのものを Cache として作成している。



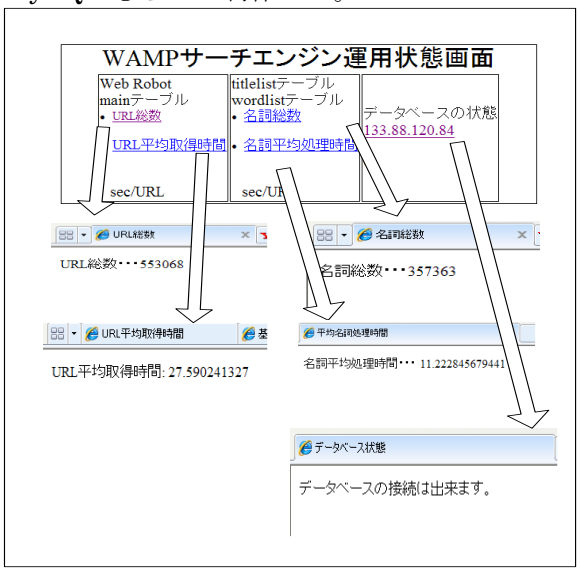
図の検索の流れを説明する。

(1) 検索画面からユーザーが検索要求
(2) キャッシュに検索要求したものがあるか問い合わせ
(3) キャッシュに検索要求したものがある場合は、それを返す→(8)にいく
(4) キャッシュに検索要求したものがない場合は、名詞テーブルに検索要求
(5) 名詞テーブルに検索キーワードがある場合は、キーワードの名詞テーブルを返す→(8)にいく
(6) キャッシュにも、名詞テーブルにも検索要求したものがない場合は、main テーブルに問い合わせ
(7) main テーブルに検索要求したものがある場合は、結果を返す
(8) 検索結果をユーザーに返答

本研究では、図で説明した主要データベースのシステムの状態を可視化したスクリプトを制作した。

3. 状態一覧スクリプト群

サーチエンジンシステム内の主要データベースの状態が分かるスクリプトを PHP と MySQL もちいて制作した。



上の図は、133.88.120.84 のデータベースの運用状態を表している。

133.88.120.84 のデータベースを可視化することにより以下の事が数字として分かる。

・URL 総数
・URL 平均取得時間
・名詞総数
・平均名詞処理時間
・データベースの接続状態

以上の可視化により、システム内のデータがどの様になっているかを知ることができる。

4. 今後の課題

制作したデータベース可視化ツールは、数字でのデータ表示しか対応していない。より分かりやすく、グラフを使った可視化が必要である。

また、制作した可視化ツールは、Cache のデータベースは対応していない。Cache のデータベースは、サーチエンジンの要でもあるため可視化は必須であると考える。

以上の 2 つの課題を解決することが、今後の課題である。