

2011年2月17日

1. 研究目的

本研究は、幸谷研究室で開発中のサーチエンジンシステムにユーザーのキーワードの履歴を保存・閲覧・判別する機能を組み込む事を目的としている。

開発中のシステムには履歴を保存する機能がなかった為、その機能を追加したいと考えた。

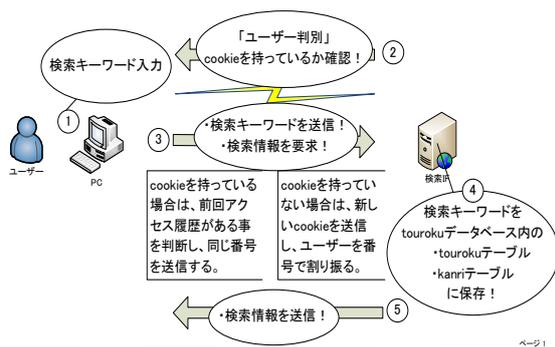
履歴にも色々あるが、本研究では、アクセスしたユーザーを判別し、入力したキーワードを保存・閲覧する機能をもったスクリプトを製作する。

2. システム概要

本研究では、PHP と MySQL を用いた検索キーワードの履歴保存だけを行うスクリプトを製作した。その機能を既存のサーチエンジン検索 IF (インターフェース) に組み込んだ。

以下の図は、その完成したシステムの流れである。

検索キーワードの履歴保存システム



システムの流れの詳細を以下に記す。

- ① ユーザーが検索キーワードを入力。
- ② 検索 IF からユーザーへ、cookie を使用した判別番号を与える。判別番号は、アクセスしたユーザー順に与える。
- ③ 検索キーワードを検索 IF に送信し、検索情報を要求する。
- ④ この際に検索キーワードを touroku データベース内の touroku テーブルと kanri テーブルへ格納する。
- ⑤ 検索情報をユーザーに送信し検索を開始する。検索システムは、関与していない為詳細は省く。

このシステムで入力されたキーワードの履歴はユーザーのブラウザで表示する事が出来、判別番号とキーワードと検索時間が表示される。

また、この履歴をユーザーがブラウザで自由に削除出来る様に削除用スクリプト(削除方法は一括削除のみ)を製作した。

3. 検索時間の検証

キーワード履歴保存システムを検索 IF に組み込んだ事で、検索時間にどの程度影響が出たのかを実験した。

以下の表は本システムの組み込み前と組み込み後の検索時間を比較したものである。

「組み込み前」

	ex time	%
Start	-	0.00%
Marker 1	0.052467	1.06%
Marker 2	4.881879	98.76%
Stop	0.008926	0.18%
total	4.943272	100.00%

「組み込み後」

	ex time	%
Start	-	0.00%
Marker 1	0.128003	2.55%
Marker 2	4.853584	96.61%
Stop	0.042278	0.84%
total	5.023865	100.00%

ex time 列の total 値は、検索時間(秒)を表す。%列の値は、検索時間の処理比率を表す。

処理は、Start (計測開始) から Marker1 (MeCab 処理), Marker1 (MeCab 処理) から Marker2 (検索結果作成), Marker2 (検索結果作成) から Stop (計測終了) までの3区間に分けている。MeCab とは、名詞のみを検索キーワードとして取り出すために導入している形態素解析エンジンである。

本システムは、Start から Marker1 までの処理区間に組み込んであり、組み込み前と処理時間を比較すると、+0.075536 秒遅くなっている事が分かる。検索時間は、+0.080593 秒遅くなっている事が分かる。

結果、組み込んだ区間の処理は倍以上遅くなっているが、検索時間を見てもほとんど変化は無い。つまり、体感速度はあまり変わらないと考えられる。

4. 今後の課題

キーワードの履歴の表示方法と削除方法に改善が必要である。

【表示】自分のブラウザには自分のキーワードのみ表示する様に改善が必要である。

【削除】必要のないキーワードだけを個別に削除出来るスクリプトを追加したい。

また、キーワード以外にもユーザーの訪問した Web ページの履歴保存が出来る様なになれば、履歴保存システムとして役に立つものになると考えられる。