

あなたのHackに検索を

Yappo

大沢 和宏

今までの主流

- 完成されたアプリを使った構築
 - Blog関連では、BulkfeedsやFeedBackなどがNamazuを用いてBlog検索サイトを構築
- 組み込み用途に特化しない
 - 一つのアプリケーションとして配布している為、組み込みずらいケースも

検索ツールに新たな波

- 今年は検索ツール元年
 - Hyper Estraier, Senna, Rastが登場
- 組み込まれる事を前提としたツール
 - 自作ツールへの組み込みも容易に
- C/Perl/Ruby/SQL(MySQL)で利用可能
 - 使い慣れた言語で, 好みの実装を

Hyper Estraier

- 類似検索
- 文書属性を自由に設定可能
- P2P対応 (予定)
- エンジンにQDBMを利用
- 付属のコマンドが充実
- C APIのみ

Senna

- 転置インデックス処理に特化
- MySQLバインディング
- 類似検索 (予定)
- C,Perl,RubyのAPIが用意されている
- 利用事例
 - はてなブックマーク
 - その他、著名なサイトでも・・・との噂

Senna + MySQL

- MySQLのFULLTEXT SEARCHを日本語対応させます
- MySQLへのパッチ
 - 本家ソースコードの一部を変更
- 利用は超簡単
 - 既存のSQLを使うだけ
- 運用にはコツも必要
 - myisamchkなどに未対応. 詳しくはドキュメントを

Rast

- XML-RPCを用いた分散機能
- 文書属性を自由に設定可能
- APR, Berkeley DBを利用
 - C API利用時にはAPRの知識も
- C, Perl, RubyのAPIが用意されている
 - Perlバインディングは開発者が別

Rast 利用事例

- tDiary-Rast tDiaryの検索
- blosxom + Rast blosxomの検索
- mod_search_rast Apacheモジュール化
- ximapd 検索ベースのimapサーバ
- gigi メールボックス検索ツール

Rast サンプル 1/2

```
1:use Rast;
```

```
2:my $rast = Rast->open("/rastdir",  
    RAST_DB_RDONLY);
```

```
3:my $result = $rast->search("検索クエリ",
```

```
4: {
```

```
5:     need_summary => 1,
```

```
6:     properties => ['title'],
```

```
7: });
```

Rast サンプル 2/2

```
8:print "Hit:" . $result->hit_count . "¥n¥n";
9:while(my $row = $result->fetch){
10:  print $row->{properties}->[0] . "¥n";
11:  print $row->{summary} . "¥n";
12:}
13:$rast->close;
```

独断で決めた比較表

	HE	Senna	S+M	Rast
installのしやすさ	?		×	
対応処理言語数			—	
利用実績	×	?		
組み込みやすさ	?	?		
分散処理		×		
分かち書き	×			
類似検索			?	×

Blogへの活用アイデア

- ツール標準の検索機能の置き換え
- 自分だけの未来検索を構築
- 関連エントリ抽出ツール
 - Hyper Estraierを使うと楽かも
- トラックバック元エントリ検索
- アクセスログ検索

さいごに

- 実は全てが和製ツール
- 泣いても笑っても、新しいツール
- Hackの道具にするもよし、これ自体をHackしてもよし