Searching in a Public Library

Some Experiences with the Search Behaviour of Patrons of a Public Library

At the public library in Waalre (NL) there is a dedicated catalogue system in operation now for 2 years. All search actions are being logged and it is interesting to analyse these in order to learn how anonymous users of the system behave.

The results of this analysis have been used to extend the functionality of the system with the aim to eventually being developed to a one stop information resource for patrons (and other users) of the catalogue system.

Many of the concepts thus developed can be "translated" to Wikipedia (and possibly other wiki's as well). The added value of this work is that it is based on the behaviour of all kind of users, from kids to seniors, from (computer)dummies to experienced information seekers.
Topics

- Introduction / Background
- SmartSearch
  - A tool assisting the user to find what he meant
  - Thesaurus
- Feedback from logs
  - To improve composition of holdings
- Helping the user
  - Integrating various sources of information
  - In finding what he intuitively is searching
- Conclusions
  - Transfer concepts to Wikipedia
Public Library Functions

- Holdings
  - Books / Magazines (and other items: CD’s etc.)
    - On loan
    - To be consulted on site
- Inter Library Loans
- Source for Information
  - Printed (books / magazines)
  - Electronic (dedicated databases, internet)
  - Librarian as intermediate
The Public Library in Waalre (NL)

Some Data

- Community population: 17,000
- Patrons: 5,000
- Holdings: 40,000
- Loans: 120,000 per annum
- Financing:
  - Municipality € 260,000 (70%)
  - Patrons € 100,000 (30%)
Similarities and Differences
between a Library Catalogue and an Encyclopaedia

• Similarities
  ▪ End-users (not editors) are very similar
  ▪ End-users are looking for information

• Differences
  ▪ Single index - Multiple Indexes
    • title, author, keyword, etc.

Normally an encyclopaedia is searched on the title of the lemma.
In a library catalogue there are multiple indexes.
This has only consequences for the technical implementation, not for
the concepts as presented here.
Design / History
of the (new) Library's Catalogue

• Front-end to existing commercial library system
  ▪ Developed from 1-1-2003 onwards
  ▪ Commissioning first version 1-7-2003

• Prerequisites
  ▪ Simplicity
  ▪ Intuitive use
  ▪ Performance
  ▪ Log everything in order to learn
    (about use, future needs, etc.)

• Ultimate goal
  ▪ Be the premier search tool for the patrons
  ▪ Weekly searches have grown from 400 to >1.600

The simplicity of Google's home page has inspired the design. There are numerous examples of commercial catalogue systems, which are too complicated for an ordinary user.

The development has been done on an old server. That forces the development process to continuously search for smart solutions, in order to provide good performance. Thus a built in warranty for scaleability is realised.
Thousands of these records, in particular the query phrase, have been browsed to understand the behaviour of the users.
Mistakes by Users

- Too many words *(each word needs to match one of the indexes)*
  - Add more, rather than delete
- Wrong index being searched
- Ordinary misspellings
- Plural / Single nouns
- Combine / Split words
Solution: SmartSearch

- Attempts to correct (or improve) the search string typed in by the user

1. Proactive in case of null results
   - Possible improvements are automatic

2. Passive in case of results returned
   - Possible extended search is suggested
Here again thinking about performance is important. The Auto Modifications are executed immediately, because there would be a null result otherwise. The Search Tip appears some time after the page is displayed, as this is processed in the background.
SmartSearch: How it works

• Plural / Single nouns
  - Default
    • Search on beginning of words
    • *house* will also find *houses*, but *houses* will not find *house*
  - SmartSearch
    • If *houses* typed in AND single form (remove last s) is an existing word, than replace houses by *house*
  - Special case: language dependent exceptions, e.g.
    • huis – huizen (Dutch)
    • haus – häuser (German)
    • duif – duiven (Dutch)
SmartSearch: How it works (2)

- Ordinary misspellings (example muullis)
  - One character
    - m*uullis
    - mu*llis, etc
  - Two characters
    - m*llis
    - m*u*iss, etc

- Join/split words
  - Example: science fiction - sciencefiction

An asterisk means that this could be any number of characters.
SmartSearch: Performance

- Display raw results (if any) immediately
- Restrict number of SQL queries on Server
  - Escape from SmartSearch procedure if result/improvement found
  - By using UNION queries where possible

SmartSearch follows a sequence to analyse the query phrase. As soon as an improvement is found that sequence is stopped and results become visible.

In particular the SQL query, investigating the occurrence of words similar to the (possibly misspelled) word(s) as typed in, should be constructed as an UNION SQL statement, in order to keep performance within reasonable limits.
Background Processing

Server

Search Frame

Result Frame

Hidden Frame

Query

Results

replace url+qs

onload

Main Page shown

JavaScript
The Result

Percentage of null results dropped from 30-35% to well below 5%.

The commissioning of SmartSearch resulted in an immediately drop of the null result rate (from 35% to < 5%). Both graphs follow the trend of increasing use in terms of searches per week.
Searching on Keywords

- Keywords to "Title Descriptions" are not consistent

- Solution:
  - Dedicated thesaurus
  - Derived from standard MS-Word thesaurus
  - Only words and synonyms that are present in "all words" table
This automatically generated thesaurus, consists of > 50,000 (valid/meaningful) records (all words in the column 'Alias' do have a reference to an existing keyword in one of the (40,000) bibliographic records.
Feedback to Librarians

- List of null results
  - Seamless link to Google to understand meaning of entered word(s)

- List of SmartSearch interferences
  - Derive obvious misspellings
  - Can be used for dedicated entries in (automatically) working thesaurus
Null-Results after SmartSearch

This is a list of null results.
The librarian can click on the original search phrase, resulting in a Google search. In this way the librarian can find out the possible meaning.

PHP is a good example, a librarian will most likely not know the meaning of this acronym, but may find out very quickly.
### Interferences by SmartSearch

#### Catalogus - Analyse Zoekresultaten

#### SmartSearch Aanpassingen

<table>
<thead>
<tr>
<th>#</th>
<th>Datum Zoekopdracht</th>
<th>Eerste</th>
<th>Laatste</th>
<th>Dorsprankelijk</th>
<th>Aangepast</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>29-7-2005</td>
<td>29-3-2004</td>
<td>royen</td>
<td>royen</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>27-7-2005</td>
<td>5-11-2004</td>
<td>hoe over leef ik</td>
<td>hoe overleef ik</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>15-7-2005</td>
<td>19-4-2004</td>
<td>mathilda</td>
<td>mathilde</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>15-7-2005</td>
<td>10-5-2004</td>
<td>turksfruit</td>
<td>turks fruit</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>29-6-2005</td>
<td>25-5-2004</td>
<td>onpersoonlijkheid</td>
<td>persoonlijkheid</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>7-6-2005</td>
<td>19-4-2004</td>
<td>gossipgirl</td>
<td>gossip girl</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1-6-2005</td>
<td>17-5-2004</td>
<td>romulus mijn vader</td>
<td>mijn vader</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>23-6-2005</td>
<td>30-6-2004</td>
<td>elfquest</td>
<td>question</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>14-5-2005</td>
<td>22-4-2004</td>
<td>starwars</td>
<td>starters</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>18-7-2005</td>
<td>9-7-2004</td>
<td>dolfwervolffje</td>
<td>dolfwervolfje</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>27-7-2005</td>
<td>29-3-2004</td>
<td>heelen royen</td>
<td>heelen royen</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>23-7-2005</td>
<td>22-5-2005</td>
<td>fernhout</td>
<td>turnhout</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>22-7-2005</td>
<td>15-7-2004</td>
<td>mcgraw</td>
<td>mcgraw</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>27-6-2005</td>
<td>24-5-2004</td>
<td>mc donald</td>
<td>mcdonald</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8-6-2005</td>
<td>13-7-2004</td>
<td>charmed</td>
<td>charme</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3-6-2005</td>
<td>21-9-2004</td>
<td>trimbaan</td>
<td>trimbach</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2-6-2005</td>
<td>25-10-2004</td>
<td>capitol</td>
<td>capitals</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>28-5-2005</td>
<td>27-2-2004</td>
<td>verrotte leven floortje bloem</td>
<td>verrotte leven floortje bloem</td>
<td></td>
</tr>
</tbody>
</table>
Provide User with alternative Information Sources

Lower bar with buttons appears after background processing. It shows only buttons of selected sites and/or Wikipedia, if that information resource has indeed an entry with in this case Amsterdam in either title or text.

In the case of Wikipedia all articles with Amsterdam somewhere in the title are shown.

<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Availability</th>
<th>Loc</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amsterdam</td>
<td>Basten, Marja</td>
<td>✓</td>
<td>AAL 1</td>
<td></td>
</tr>
<tr>
<td>Amsterdam</td>
<td>Weidemann, Siggi</td>
<td>✓</td>
<td>AAL V</td>
<td></td>
</tr>
<tr>
<td>Amsterdam</td>
<td>Basten, Marja</td>
<td>✓</td>
<td>WAA 1</td>
<td></td>
</tr>
<tr>
<td>Amsterdam</td>
<td>MacEwan, Ian</td>
<td>✗</td>
<td>AAL V</td>
<td></td>
</tr>
<tr>
<td>Amsterdam in de Gouden Eeuw</td>
<td>Emeis, M.G.</td>
<td>✓</td>
<td>AAL 1</td>
<td></td>
</tr>
<tr>
<td>Amsterdam in de Gouden Eeuw</td>
<td>Emeis, M.G.</td>
<td>✓</td>
<td>WAA 1</td>
<td></td>
</tr>
<tr>
<td>Amsterdam Zuidoost</td>
<td>Osch, Yvonne van</td>
<td>✓</td>
<td>AAL 1</td>
<td></td>
</tr>
<tr>
<td>Amsterdam</td>
<td>Osch, Yvonne van</td>
<td>✓</td>
<td>WAA 1</td>
<td></td>
</tr>
<tr>
<td>Maleis</td>
<td>Asterisk®, Amsterdam</td>
<td>✓</td>
<td>AAL V</td>
<td></td>
</tr>
<tr>
<td>Boeda</td>
<td>Limburg, Inger</td>
<td>✗</td>
<td>AAL V</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asterisk®, Amsterdam</td>
<td>✓</td>
<td>AAL V</td>
<td></td>
</tr>
</tbody>
</table>

1 2 3 4 5 6 7 8 9 10 11 12 13 14 Next
Crucial for the performance of the background process is the existence of an "all words" table for every external information source.
"All Words" Table

- Where clauses in SQL statements
  - *Like* clauses are "expensive"
  - *Between* "cheap"
- Requires an "all words table"

<table>
<thead>
<tr>
<th>cur_id</th>
<th>cur_title</th>
<th>cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>109</td>
<td>Algemene_Centrale_Der_Liberale_Vakbonden_Van_België</td>
<td>0</td>
</tr>
<tr>
<td>110</td>
<td>Centrale_Der_Liberale_Vakbonden_Van_België</td>
<td>1</td>
</tr>
<tr>
<td>111</td>
<td>Der_Liberale_Vakbonden_Van_België</td>
<td>2</td>
</tr>
<tr>
<td>112</td>
<td>Liberale_Vakbonden_Van_België</td>
<td>3</td>
</tr>
<tr>
<td>113</td>
<td>Vakbonden_Van_België</td>
<td>4</td>
</tr>
<tr>
<td>114</td>
<td>Aboe_Hidal</td>
<td>0</td>
</tr>
<tr>
<td>115</td>
<td>Midal</td>
<td>1</td>
</tr>
<tr>
<td>116</td>
<td>Arabische_Liga</td>
<td>0</td>
</tr>
<tr>
<td>117</td>
<td>Liga</td>
<td>1</td>
</tr>
<tr>
<td>118</td>
<td>Afrikaans_Nationaal_Congres</td>
<td>0</td>
</tr>
<tr>
<td>119</td>
<td>Nationaal_Congres</td>
<td>1</td>
</tr>
<tr>
<td>120</td>
<td>Congres</td>
<td>2</td>
</tr>
<tr>
<td>121</td>
<td>Anti-Revolutionaire_Partij</td>
<td>0</td>
</tr>
<tr>
<td>122</td>
<td>Partij</td>
<td>1</td>
</tr>
<tr>
<td>123</td>
<td>Afro-Aziatische_talen</td>
<td>0</td>
</tr>
<tr>
<td>124</td>
<td>talen</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>Anatolische_talen</td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>talen</td>
<td></td>
</tr>
<tr>
<td>127</td>
<td>Archosauria</td>
<td></td>
</tr>
</tbody>
</table>

A (Wikipedia) table as shown here, significantly contributes to the performance of the background process.
Between ... And ...z SQL clauses are much more efficient than ordinary Like clauses in the SQL statement.
If the User is not really knowing what to look for

- The patron’s dilemma
  - I've read a book by ……, did like it (or dislike)
  - I’m looking for something similar

- What do you suggest?
Simple Cross-references
similar to e.g. Amazon buying suggestions

The pop-up screen is shown when the user clicks on an yellow square.
Complex Cross-references

Same as Simple Cross-References
But improve relevance ranking by also using the patron's own loan history to determine relevance.

The upper pane shows the loan history of an user.
The lower pane shows the suggestions, taking into account the loan history of all patrons, but the sorting is heavily influenced by the user's own loan history.
The ultimate Goal of a free Source of Knowledge

- Current wiki projects are predominantly focusing on **active users** (contributors)

- Next phase:
  - Wikipedia to be a natural info source for **passive users**
  - Passive users will behave very similar as library patrons

Will make all kind of mistakes
Have limited skills to define proper search strings
Conclusions and Recommendations (1)

- SmartSearch
  - Concepts can equally be used in e.g. *Wikipedia*
  - Algorithms for Dutch need to be converted to other languages
  - Requires another mechanism to start a new article

- Feedback to Librarians
  - Can be used as input to "Requested Articles"
  - Can be used to build meaningful Redirects
Conclusions and Recommendations (2)

- Integration of other Information Resources with Wikipedia
  - Could be made 'on the fly' if an 'all words' (in titles) table in the on-line database could be *queried*

- Cross-references (far future?)
  - Can be used to provide suggestions for "*further reading*"
Catalogue on the Internet

www.obwaalre.nl
www.obwaalre.nl/cat/default.asp?lang=en

Home page and catalogue. Most of the pages can be displayed in English.

Contact author on:
w:nl:gebruiker:RonaldB