

# Advanced Knowledge Base Concepts: Optimizing the Self-Service Experience

Doug Warner  
Head of Research  
RightNow Technologies



# Overview

- Theory
  - Spidering and Indexing
  - Search Query
- Application
  - Configuration
  - Analysis
  - Tuning



# Background

- 80% of RightNow users never use searching
  - Other sessions describe how to optimize our product for this group
- Of the remaining 20% who do search:
  - 95% view only the first page of results
  - Each search session averages 2.7 searches
- Optimizing search performance is difficult

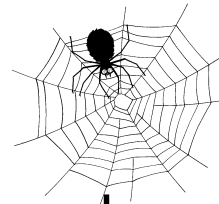


# Searching Theory

- Spiders travel links within the “web” of documents
  - Root document
  - Indexes each page found
  - RightNow Feature: single doc spider
- Indexes store the document contents
  - Like a book index: maps word to page
  - Index must contain everything searchable



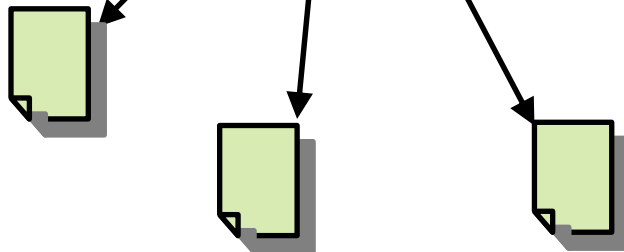
# Spidering



The [North American Continent](#) consists of the countries: the [United States of America](#), [Canada](#), [Mexico](#).

[Four score](#) and seven years ago our fathers brought forth on [this continent](#), conceived in [Liberty](#), and

Liberty:  
The condition of being free from restriction or control.



# Index

## Index

Word	Location
liberty	
mexico	
north	
restriction	
score	
seven	
states	
united	
years	

[Four score](#) and seven years ago our fathers brought forth on [this continent](#), conceived in [Liberty](#), and


Score:  
A group of 20 items. Hence, four score is  $4 \times 20$ , or 80.

Liberty:  
The condition of being free from restriction or control.

The [North American Continent](#) consists of the countries: the [United States of America](#), [Canada](#), [Mexico](#),



# Word Stemming

- Standard Natural Language Processing technique
  - Unique algorithm for each language
- Takes a word to its base form:  
walked, walks, walking  walk
- Generalizes searches (exact matches not considered)



# Stop Words

- Standard indexing technique
  - Removes common words from index
    - o AN, ARE, BE, CAN, HAVE, IS, ...
  - Increases search performance
  - RightNow Feature: multiple editable stop word lists
    - o Incidents
    - o Answers



# Search Query

- User queries compared against index
  - Stemming
  - Stop words
- Scores generated from amount of match
  - Word/phrase frequencies
    - o 10 repetitions score higher than 1 occurrence
  - Word/phrase locations
    - o Title words score higher than body words



# Search Query

- AND/OR decisions done at query time
- Number of results depends on scoring algorithm
- Result ranking depends on scoring
  - Single word repetitions more important than multi-word phrase match?
- RightNow Feature: views modify search results, manipulating visibility with word-level granularity



# Application: Configuration

- Config settings:

- Word bias (common.msg)

o SRCH_KEY_WEIGHT	50	Keywords
o SRCH_PROD_WEIGHT	50	Product Words
o SRCH_CAT_WEIGHT	50	Category Words
o SRCH_SUBJ_WEIGHT	45	Subject/Title Words
o SRCH_DESC_WEIGHT	30	Question Words
o SRCH_BODY_WEIGHT	4	Answer Words
o SRCH_ATTACH_WEIGHT	4	File-Attach. Words

- AND/OR (rnw\_ui.msg)

- o ANS\_AND\_SEARCH



# Application: Configuration

- Word lists
  - Stop words
    - o exclude\_answers.txt            End-user stop words
      - Short list; small number of documents allows a trade of speed for more results
    - o exclude\_incidents.txt        Incident stop words
      - Long list; large number of documents and trainable users allows a trade of certain results for more speed
  - Aliases
    - o aliases.txt                    Always-On Search Thesaurus
  - Related lists
    - o thesaurus.txt                Only used for Similar Search
    - o userdic.tlx                    Only used for spell checking



# Application: Configuration

- Management and Configuration
  - Views
    - o End-User answer view
    - o Admin-User answer view(s)
    - o Admin-User incident view(s)
    - o Consider the matched weight column
  - Topic words
    - o Customize Answer search results by phrase
      - Results appear before normal list
      - Search-independent option
      - “Fix-at-top on steroids”



## Answer Console

Quick Search ▾  Refresh Search

New ▾ Edit ▾ Delete ▾ Print Preview

Viewing Answer\_Find

Records Found: 164

	Met	ID	Summary	Status	Score	Weight	8.0 Prep St	Content O	Access Le	Date Updat
1	1355	1355	When searching for incidents and answer:	Public	1562	100	Minor Edit		Everyone	07/31/2006 10:00
2	1346	1346	What is the difference between weight, sc	Public	658	98	Minor Edit		Everyone	07/31/2006 09:00
3	409	409	Moving a public answer up or down in the	Public	1506	98	Minor Edit		Everyone	07/31/2006 11:00
4	575	575	In 5.x, Customizing the Answers List page	Public	2571	81	Check Conter		Everyone	07/28/2006 09:00
5	1767	1771	Affecting the weight used with search res	New to 6.0	6195	81	Minor Edit		Everyone	07/28/2006 09:00
6	1991	1995	Using a list of keywords for all answers	New to 7.0	567	80	Minor Edit		Everyone	08/18/2006 09:00
7	936	936	Improving Smart Assistant™ to suggest mc	Public	4406	55	Minor Edit		Everyone	07/28/2006 11:00
8	2048	2052	What is the Information Gap report?	New to 6.0	340	54			Everyone	07/27/2006 10:00
9	2431	2487	Reducing the number of answers returned	New to 6.0	269	54	Minor Edit		Everyone	08/01/2006 09:00
10	2254	2258	Always having an answer be first in the s	New to 7.0	96	54	Verify on Rel		Everyone	08/18/2006 09:00
11	1336	1336	How do I delete a criteria or output column	Public	54	54			Everyone	07/27/2006 10:00
12	1092	1092	In 5.x, how can I customize the views for t	Only in 5.x	481	53	Complete		Everyone	07/27/2006 09:00



# Support Console

Quick Search ▾ | Incidents by Keywords - Best Match | Refresh | Search | New ▾ | Edit ▾ | Delete ▾ | Print ▾ | Forward | Propose

## Viewing Incidents by Keywords - Best Match

Records Found: 2313

	Reference	Subject	Sever	Date Cre	Organizatio	Stat	Support	Version	Weight
1	060613-000	VER: Views: View search functionality for 'image file attachr	2 - Majc	06/13/2006	Linksys - Sub:	Unres	Preferred	7.5.4.0-b358h-mysql	100
2	060628-000	7.5.3.0 - Unable to access file manager files for other interfac	3 - Som	06/28/2006	NC Interactive	Updat	Preferred	7.5.3.0-b327h-mysql	92
3	060808-000	Staff unable to use the "save" button in an FAQ	3 - Som	08/08/2006	DivX, Inc.	Waitir	Preferred	7.5.4.0-b358h-mysql	92
4	060531-000	Setting up ID's for a customer	3 - Som	05/31/2006	Dun & Bradstr	Solve	Preferred	7.5.4.0-b358h-mysql	91
5	060622-000	Missing FAQs (Public FAQs not Dispaled on the End User Par	4 - No F	06/22/2006	Reuters PLC	Solve	Preferred	7.5.4.0-b358h-mysql	91

### Search

<b>Status Type</b> Unresolved Solved Waiting Unspecified	<b>Group</b> Tier 1 Tier 2 Tier 3 Tune-Up Tech Support - APAC Upgrade Specialist Voice Unspecified	<b>Product Hierarchy</b> <input checked="" type="checkbox"/> Administration Cor <input checked="" type="checkbox"/> Analytics and Rep <input checked="" type="checkbox"/> RightNow Service <input checked="" type="checkbox"/> RightNow Live <input checked="" type="checkbox"/> RightNow Locator <input checked="" type="checkbox"/> RightNow Marketin	<b>Category Hierarchy</b> <input checked="" type="checkbox"/> Getting Started <input checked="" type="checkbox"/> Configuration <input checked="" type="checkbox"/> Best Practices <input checked="" type="checkbox"/> Page Descriptions <input checked="" type="checkbox"/> Installations and L <input checked="" type="checkbox"/> Troubleshooting	<b>Dormant</b> Yes No Unspecified	<b>Status</b> Results Rejected MKT-Unresc MKT-Waiting MKT-Solved Solved- Cus Non-Reprod Unspecified
--	--	---	--	--	---

**Date Time Search** \* Date Created ▾ 5/1/2006 12:00 AM ▾  relative to 8/18/2006 12:00 AM ▾  relative

**Text Search** \* Summary/Thread ▾ handy views

**Order By** Weight - desc ▾ -- ▾ **Limit to** 500 rows **Reset** **Search** **Cancel**





**Services**

**BUSINESS**

- Voice Services
- Data Services
- Network Solutions
- Solutions by Business type
- Business ADSL

**HOME**

- Billing Information
- Mobile Phones
- Modem Services
- Voice Services
- Support and Advice
- Home ADSL

Save up to 30% on phones and office equipment

**Online Services**

- My Bill
- My Account
- My Questions

Products Coming Soon

Clearance Offers

What's Hottest

» Online Assistance

Support Home **Answers** Ask a Question My Stuff Login Help

Search Browse

Product **i** Keyword Search Search Tips  
All All billing Search

Category **i** Search By Sort By  
All Phrases Default Sort Powered by RIGHT NOW

Other Suggested Searches: [online billing?](#) [new account](#) [account](#)

**Recommended Documents**

- 1 [What is ONE-BILL?](#)  
New Offer! Save time and money! Enroll in One-Bill today! Click here for details...

**9 Answers Found**

	Summary	Score
1	Paying Your Bill Online	7
2	Paying Your Bill with a Credit or Debit Card	6
3	What is a credit limit?	141
4	Smartguide Sample	35
5	What is ONE-BILL?	15
6	Type of ADSL Modem	56
7	Questions about Long Distance	28
8	What is the cost of a Cable Modem for Internet Access?	6
9	Step by Step Guide for Modem Trouble shooting	2



# Application: Analysis

- Keyword Searches report
  - Frequent searches (important content)
  - Searches with no answers (missing content)
  - Searches with too many answers (configuration and tuning needed)
- Custom report:
  - Answer ID, Summary, Keywords (potential aliases)



# Keyword Searches

## Phrase Stem

Phrase Stem	Search Count	Answers
68	6	0
1351	6	0
GEXPAND	6	0
CASTL	6	0
OHTANI@RIGHTNOW.MITSUI.CO.JP JALCARD	4	0
768	4	0
MY_SEC_COMPANI	4	0
CNAME	9	1
1772	8	1
IE7	8	1
ENVIRONMENT	8	1
RNTINFO	8	1
ARCHITECTUR	7	1
STREAM	6	1
SAWMIL	6	1



# Keyword Searches

## Phrase Stem

Phrase Stem	Search Count	Answers
REGIST STAFF RIGHTNOW SUPPORT	5	1165
OBJECT DOESN SUPPORT PROPERTI METHOD	10	896
LANGUAG RIGHTNOW WRITTEN	4	886
CUSTOM END USER PAGE	6	877
HTML CONFIGUR	4	875
SET UP ORGAN	8	863
AD DATE FIELD SEARCH SUPPORT CONSOL	4	863
OBJECT DOE SUPPORT	4	818
SET PRIORITI LEVEL HIGH IMPORT INCID	4	808
FAIL CONFIGUR CONNECT UNKNOWN INCORRECT TIME ZONE PST8PDT	6	802
END USER PAGE	4	799



# Application: Tuning

- Content
  - Keywords
    - o Optional—try aliases.txt wordlist first
    - o Highly weighted (SRCH\_KEY\_WEIGHT 50)
    - o Phrases separated by comma
    - o Not displayed: consider repetition, competitor equivalents, etc.



# Application: Tuning

- Content
  - Consistency
    - o Write well, follow internal authoring guidelines
    - o Put important terms in important locations (title, keywords)
    - o Make important location content distinct (web pages: don't always use the same <TITLE>)
    - o Avoid acronyms, ambiguous phrases, unusual punctuation
    - o See <http://www.WebWritingThatWorks.com>



# Summary

- Search tuning is difficult
- Search results are in a “race” to the top
  - Each reordering of results may have wide implications
- You can’t find things not in the index, but you can’t find things if the index is too big!
- Help the system by using easy to parse/differentiate text



# Questions?

- Search engine differences: Google, Yahoo, MSN, RightNow
- Reports
- Configuration
- Website Indexing
- Only 20% of users search?

