Lunch and Learn: visionSearch

July 2016

Agenda

- visionSearch Overview
  - What is it?
  - Features
  - Competition
  - Positioning

- visionSearch In-depth look
  - Indexing
  - Search Results

- visionSearch Demo

- Questions and Answers

- Resources

visionSearch Overview

visionSearch: What is It?

- A search option that Vision developed in-house using a popular open-source search library (Solr)
  - Customize Search Results
  - Configure and Schedule Indexing
- A search replacement for Searchblox
- A search alternative to Google Search (CSE)

- Included in all new site implementations as of February 1, 2016

- Compatible with intranets
  - Does not require an internet connection (Google does)

visionSearch: Key Features

- Basic Search Options
  - Write Boolean searches without having to know Boolean logic
    - For example, search with all of the words, with the exact phrase, with at least one of the words or without the words
  - Sort results by relevance, last modified, title, or content type
  - View result attributes, such as content type (Page, News, PDF, etc.), file size, and last modified date

- Customization of Search Results
  - Backend: Add search synonyms to make it easier for users to find the most relevant result
  - Backend: Promote frequently accessed or highly relevant pages within search results
  - Frontend: Filter results, displaying only specific file types (Adobe PDFs, Word DOCs, etc.)
  - Frontend: Filter results by content type (for example, show only News and RFPs)

visionSearch: Key Features

- Advanced Search Features
  - Intelligent indexing
  - Auto-suggest makes dynamic search recommendations, as the query is typed in the search field
  - Fuzzy logic and Boolean keyword searches

- Configuration Options
  - Ability to define the scope of the search index (e.g., one department or entire site)
  - Intranet compatibility
  - Configure and Schedule Indexing
  - Index and include results from some third-party sites

visionSearch: New Features

- Member-only Access (as of June 2016)
  - Intranet websites hosted by Vision may enable "member-only" access for all front-end pages instead of enabling front-end protections
    - Clients prefer this route when different departments use different member groups
    - This allows the organization to limit access to certain pages for that department
  - visionSearch is now capable of indexing member-only pages
    - Turned off by default
    - A client must specifically request this functionality before the Production team activates the feature

visionSearch: Competition

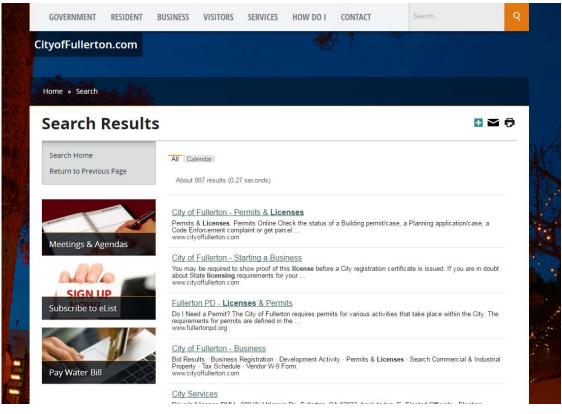| Competitor	Grade	Comments
Vision	A	• Basic & Advanced options • Sort results by relevance, last modified, content type, and title • Filter by content type on basic search page • Better styling • Fine tune results with synonyms or key words
CivicPlus	A-	• Basic & Advanced options • Filter by content type (one type at a time) on advanced search page • No sorting options • Site-specific search suggestions
Granicus-Civica	B+	• Basic only • Filter by Calendar, PDF • Site-specific search suggestions
Revize	C	• Basic only • Results include external ads
CivicLive	C	• Basic only • Results include external ads

CivicPlus

City of Hutchinson, KS
Standard and Advanced options
Advanced – simplified Boolean and filter by Content Type
No sorting capabilities by date or relevance

Granicus-Civica

Fullerton, CA
Basic search only
Filter All or Calendar, PDF
Site-specific search suggestions on some sites

Revize

North Las Vegas
Powered by Google Custom Search
Includes ads in their search results

CivicLive

North Little Rock
Basic sorting by Relevance and Date
Includes ads in their search results

visionSearch: Positioning

- Robust search is imperative for a customer-centric site
  - Research has shown that search is one of the top 3 website actions taken on government websites
  - 90% of users do not go past the first page of search results

- Fine-tuning search improves user experience and results
  - Set up search synonyms to capture all possible phrases your site visitors may use
  - Use keywords to promote specific pages to the top of the search results

- Customers get a better search experience and can quickly find what they want
  - Sort results to quickly find key information
  - Filter results by content type to narrow results
  - Access advanced search options with a simple click

visionSearch: In-depth Look

VISION

Indexing

- In order to display results, we need to be able to index the content
  - Can be configured to index either specific pages (e.g., department section of a website) or the entire site

- Crawl Depth
  - Crawl depth is the extent to which a search engine indexes pages within a website (home page crawl depth = 0; Pages in the same site that are linked directly (with one click) from within the home page have a crawl depth of 1)
  - Default is 7
  - If the depth is 7, visionSearch will index the additional levels pages when we run visionSearch the next time
  - It may take more time to index all levels' pages

- Frequency
  - Scheduled weekly by default

- Clear
  - Scheduled weekly by default
  - This will clear all indexed results. After the "clear," we will re-index the site again.

Search Results

- visionSearch results are based on Solr, which uses the Porter Stemming Algorithm

- Porter Stemming Algorithm
  - Stemming is finding the basic form of a word
  - The Porter Stemming Algorithm is a transforming algorithm for English language that reduces any of the forms of a word such as "walks, walking, walked", to its elemental root e.g., "walk"
  - Porter is rules based and does not need a dictionary
  - Complexity of the English language means that a search may return some results that are not expected
  - Pros outweigh the cons

visionSearch **synonyms** and **page promotion** options facilitate displaying most relevant results at the top of a search

visionSearch Demo

Demo

Demonstration

Questions and Answers

Questions and Answers

- I understand VisionSearch can search other information that is "Google-able"; but can you explain this in more detail?

- What about document management system searches (DMS)? Laserfiche, SIRE, others? Can VisionSearch pull information from these databases and be included with a client's search?   What are the limitations here?

- Also along with the above, please list or show any examples that we already have in place for searching 3rd party repositories

Questions and Answers

- visionSearch does include the option to include results from other sites, including third parties

- In order to include results from a third party, the following criteria must be met
  - URL must be available without authentication or other interface preventing indexing
  - Client may need to add visionSearch as an "approved" search engine to allow for indexing (i.e., not blocked by robots.txt)
  - Results formatting must be compatible (some formatting may not be compatible with visionSearch)
  - Content must be available for indexing
    - We can include PDFs and content from other systems if we are able to index that content

Laserfiche with visionSearch Integration

- visionSearch can display Laserfiche documents if they are able to be indexed

- We've offered visionSearch-Laserfiche integration to certain clients, on a case-by-case basis (e.g., St. Lous Park, MN; Bozeman, MT)

- Depends on the system and the way it is set up

- Client would need to:
  - Open the Laserfiche application portal, so that visionSearch can index the documents from the Laserfiche application
  - For better performance, the maximum document size visionSearch can index is 50 MB

- In addition to an implementation fee, a recurring fee would also apply. Recurring fees may vary, depending on:
  - Total number and total size of documents to index in Laserfiche
  - Bandwidth required to index (when we index, we have to download all documents)
  - Resources and time required to index

Questions to Ask when Evaluating visionSearch Integration

- What third-party software are they using?

- Is it currently being indexed by another search engine (like Google, Bing, Yahoo)?
  - If these search engines were able to index their documents, then most likely the answer is "Yes, visionSearch can index them too"
  - As long as "other" search engines are not restricted in the robots.txt file

- Does the third-party software block bots/web-crawlers from indexing their site or files?

- Does the third-party software provide an indexed sitemap that search engines can use to traverse through their documents?

Questions and Answers

- Do we use the database of the site at all in building the search index?
    - No, visionSearch indexes each page based on its HTML

- How long does it take to re-index a site completely?
    - It depends on how much content the site has
    - Typically, it takes 1 hour to 3 hours
    - If there are multiple sites setup into one visionSearch instance, it will take even longer (could be 12 hours or longer)

Questions and Answers

- How do we handle updated links?
  - Generally, we will schedule a weekly "clear." This will clear all indexed results. Then we will re-index it again.
  - The other way is to setup the document age (e.g., 7 days), which means once the document has been indexed for 7 days, it will be updated next time

- Do the search results use the navids or the friendly URLs?
  - visionSearch always uses the URL in the page's HTML (can't tell what's a navid or friendly URL)

Resources

- Additional Resources Available on Insight
    - This Presentation
    - Basic Description
    - User Guides
    - Configuration Documentation
    - Stemming Logic Links