visionSearch Integration

# visionSearch

1. visionSearch can index pages & documents from an open portal.
2. visionSearch can index below types of documents:

HTML, PowerPoint, PDF, RTF, Text, Word, Excel, Open Office.

If client wants to index other types, please contact product team or software development team.

1. visionSearch support Content Type filter.
2. visionSearch support synonym.
3. Each visionSearch instance can index multiple different sites.
4. By default, we will schedule visionSearch index every day and clear every week. After cleared, all existing index will be removed, and visionSearch will reindex everything.

If don’t schedule clear, by default, existing indexed documents will be re-indexed after 7 days.

# CMS6 Integration

1. We can configure multiple visionSearch instances in one CMS6, if client has paid.
2. In CMS6, Each search widget can only connect to one visionSearch instance.
3. We can’t display the search results from two different visionSearch instances in one search widget
4. In CMS6, we can use Meta Keyword and Meta Description of page to increase search ranking.

# Third Party Requirements

1. The third party application portal needs to be open to visionSearch without any login.
2. The robots.txt of third party application should not block visionSearch.
3. Documents in the third party should not be encrypted.
4. For better performance:
	1. The maximum document size visionSearch can index is 50 MB.
	2. The maximum folder level (crawl depth) is 7.

# Third Party Information Needed

1. Index portal URL.
2. Index scope (include).
	1. Option 1: Index all pages/documents
	2. Option 2: Index some pages/documents.

Please provide the URL format or document type format or folder structure, which client wants to index.

1. Index scope (exclude)

If client want to exclude some pages/documents for visionSearch, please provide these URL format or document types.

1. Index document types (if not only pages)

Please give all document types if applicable which client wants to index, and give a sample URL for each type.

Note: if client has multiple sites which need to be indexed, please ask above information for each site.

# Third Party Application Integration Case Study

## Laserfiche (Client: Bozeman, MT, http://dev7.visioninternet.com/BozemanMT6/ )

1. Laserfiche Portal: <http://weblink.bozeman.net/WebLink8/>

visionSearch Root Page: <http://weblink.bozeman.net/WebLink8/Browse.aspx?startid=26>

1. Robots.txt URL: <http://weblink.bozeman.net/robots.txt>

Don’t have robots.txt, which means visionSearch can index.

1. Index Scope (include)

Only 3 folders:

* \City Commission
* \Economic Development
* \GIS
1. Index Scope (exclude): N/A
2. Document Types
	1. TIF - http://weblink.bozeman.net/WebLink8/DocView.aspx?id=55604
	2. PDF - http://weblink.bozeman.net/WebLink8/0/doc/76359/Electronic.aspx
	3. XLSX - <http://weblink.bozeman.net/WebLink8/0/doc/78856/Electronic.aspx>
	4. DOC/DOCX may be future.

Note:

Bozeman’s Laserfiche has OCR, i.e. <http://weblink.bozeman.net/WebLink8/DocView.aspx?id=55448&dbid=0>, since these OCR are image, not PDF version, visionSearch doesn’t support.

## OCR PDF (City of Torrance)

1. <http://arch.torranceca.gov/archivedirectories/>
2. <http://arch.torranceca.gov>
3. Robots.txt: http://torranceca.gov/robots.txt

## Granics

Will add in the future if we have.