

JChem Search for SharePoint

Tamas Pelcz



ChemAxon
Solutions for Cheminformatics

JChem Search for SharePoint

- What it is?
 - Extension to Microsoft SharePoint Search and FAST Search to be able to index and query chemical structures
- Why with Microsoft SharePoint?
 - Most widely used enterprise search technology.
- Why together with SharePoint search?
 - Less administrative effort
 - Crawl only once
 - Leverage on SharePoint features
- Possible to use without SharePoint?
 - Possibly, but have not tried yet.
 - Desktop Search, Microsoft Search Server.

Where are we? - Indexing

- JChem for SharePoint content
 - List, Blog, Discussion Board, Wiki
- Structure files: MOL, SDF, CDX, SKC, MRV
- Extracted from: IUPAC Name, SMILES
- Configurable words and synonyms:
 - Brand names: Aspirin, Viagra
 - Corporate ID: PHARMA-00000211
- OLE Objects
 - Marvin
 - ISIS Draw, Symyx Draw
 - ChemDraw
- Excel chemistry Add-In
 - JChem for Excel

Where are we going? - Indexing

- Pluggable, extensible architecture
 - Structure index storage
 - Keyword or text parsers
 - Custom entity extraction
 - Custom connector
 - Image – structure recognizer
- Custom connector to external applications
 - ELNs
- Configurable words and synonyms from database
- Excel chemistry Add-In
 - Exact cell location of the hit
 - ISIS for Excel and maybe others

Where are we? - Querying

- Web service processing queries and returning hits
 - Hides the complexity of merging structure hits from structure index and keyword hits from SharePoint
- Paging
- Refiners
 - Structure refiner
 - Content (file) type
 - Structure type (in what format the structure was found)
- Web based UI
 - Customs
 - Built in search integration
- Windows Forms sample application

Go back

1-5 of 58 results

1 2 3 4 5 6 7 8 9 10 11 12

Click on a refiner to filter search results by that category

Result Type

Any Result Type

- Adobe PDF (1)
- Chemical File (2)
 - Marvin Document (1)
 - Structure Data (1)
- Office (22)
 - Word (18)
 - Excel (4)
- Text (1)
- Webpage (28)



Molecule Format

Any Molecule Format

- Binary (8)
 - Chemical File Molecule (4)
 - JChem for Excel Molecule (2)
 - Binary List Molecule (2)
- OLE Object (6)
 - ChemDraw OLE Object (2)
 - Marvin OLE Object (2)
 - Isis OLE Object (2)
- Textual (109)
 - Iupac Name (75)
 - Smiles (24)
 - Custom Identifier (10)


Structure


Any Structure

Test Blog


There are no items in this list. ... This one of my first posts. I created it to test molecule indexing, which should find the benzene here. ... Welcome to SharePoint Blogs. Use this space to provide a ...
<http://test.sharepoint.chemaxon.com/IndexingTest/TestBlog>


 **Iupac name**
Post about **naphthalene**... I created it to test molecule indexing, which should find the **naphthalene** here.

 **Iupac name**
Post about **anthracene**... I created it to test molecule indexing, which should find the **anthracene** here.

Test List


<http://test.sharepoint.chemaxon.com/IndexingTest/Lists/Test List/AllItems.aspx>


 **Custom identifier**
SMDL-0000002
Iupac name
Naphthalene
Smiles
C1=CC2=CC=CC=C2C=C1

 **Custom identifier**
SMDL-0000003
Iupac name
Anthracene
Smiles
C1=CC2=CC3=CC=CC=C3C=C2C=C1

Test Wiki - Recent Changes

WebPageLibrary ... Site Actions ... This page location is: ... Main ... IndexingTest ... Home ... Recently Modified ... Wiki Page - Benzene ... Wiki Page - Naphthalene ... Wiki Page - Anthracene ... Sites ... Test Blog ... Libraries ... Test Documents ... Lists ... Recycle ...
<http://test.sharepoint.chemaxon.com/IndexingTest/Test Wiki/Forms/RecentChanges.aspx>

 **Iupac name**
Wiki Page - **Naphthalene**

 **Iupac name**
Wiki Page - **Anthracene**

Test Wiki - Created By Me

WebPageLibrary ... Site Actions ... This page location is: ... Main ... IndexingTest ... Home ... Recently Modified ... Wiki Page - Benzene ... Wiki Page - Naphthalene ... Wiki Page - Anthracene ... Sites ... Test Blog ... Libraries ... Test Documents ... Lists ... Recycle ...
<http://test.sharepoint.chemaxon.com/IndexingTest/Test Wiki/Forms/CreatedByMe.aspx>

Where are we going? - Querying

- Structural property refiner
 - Pre-calculated on indexing
 - Using sliders
- Structure only search
- Visual Refiners – Pie chart
- Document Preview

Architecture

- Native IFilter proxy connects to external managed indexing service
 - Similar to Microsoft Business Intelligence Connector
 - Robust, scalable, fail proof – does not affect normal indexing if something goes wrong
- FAST Integration in Content Processing Pipeline
- Query Web Service
 - Later SharePoint service application

Demonstration

- Search on small sample content, containing all possible recognizable file types, content types
- Search on indexed list of drugs from Wikipedia

TODO

- Benchmarks on indexing and querying
- Installers
 - Separate for components
 - Single server bundle
- Configuration tools
 - Desktop, SharePoint Central Admin
- Large scale deployment testing

Team

- Tibor Erdélyi

- József Dávid



Get involved

- Demo server: <http://sharepoint.chemaxon.com/> - June
- Installer – June