Kos and Himmler *Journal of Cheminformatics* 2010, **2**(Suppl 1):P1
http://www.jcheminf.com/content/2/S1/P1

Journal of Cheminformatics

POSTER PRESENTATION Open Access

CWM global search - an internet search engine for the chemist

Alexander Kos*, H-J Himmler

From 5th German Conference on Cheminformatics: 23. CIC-Workshop
Goslar, Germany. 8-10 November 2009

The Internet is a rich source of data and information for chemist. There are numerous multidisciplinary databases available for free on the Internet. Some examples of such data repositories are: PubChem, ChemSpider, eMolecules, Drugbank, KEGG, NIST, ChemSynthesis, PharmGKB, Free patents online ...

It should be obvious that an end user is a) not aware of all the resources, and b) has not the time to learn every user interface and is unable to search over all of them. We provide CWM Global Search as an application that enables to search by structure, CAS Registry Number and free text over all these sources. Presently CWM Global Search performs searches in 30 databases and search engines accessing more than 100 million pages that associate data with structures.

The user can submit a single query structure or several using SDFiles. In addition to molecule searches CWM Global Search also allows to submit reaction queries. In that case several single molecule searches are performed for the reactants, reagents and products. This makes it easy to find commercial suppliers and other synthesis relevant information such as safety sheets in one query.

Searching is technically less problematic than providing the answers in a digestible way for the user. Our first approach is to provide profiles for searching. You can choose "Availability" if you are interested to find a commercial supplier, or "Biology" if you are looking for biological effects. The second help comes when we display the summary of the results. You get a table with hyperlinks color coded by topics. If the result page of doing a search contains a link to an MSDS, the topic 'Safety' is highlighted. With profiles you limit your search to certain sources, and with topics your answers will be ordered.

CWM Global Search is not the application for exact searches like "give me the melting point of anthracene". You will find the melting point on many pages, and the topic "Physical Property" might help, but, in this case the link to Wikipedia gives the result quickest. Internet pages provide us the data in unordered fashion and finding the exact answers is time consuming. In some case like commercial suppliers we check against an internal list if the page really displays a supplier, or, if for instance PubChem has only a reference to ChemSpider, and ChemSpider references again just PubChem, but nowhere you will find a supplier. We also have to consider that many providers of the resources would not allow us to extract data directly without leading the user to their Internet pages. The nature of the results is fuzzy in CWM Global Search. This is an advantage if you look for instance for biological effects, which can be many, and/or if you want to learn why a compound could be important.

We generate both InChI names and keys for the query structure and afterwards perform fast text searches in the various databases or search engines. In addition we also generate Smiles. Since prior to the existence of standard InChI's (released 2009) the various database providers used different settings to generate the InChI's stored in their database, we use multiple settings when generating an InChI identifier used for a structure search in CWM Global Search. This way we greatly maximize the chance to find an InChI independent of the settings used by the database provider.

Published: 4 May 2010

doi:10.1186/1758-2946-2-S1-P1
Cite this article as: Kos and Himmler: CWM global search - an internet search engine for the chemist. *Journal of Cheminformatics* 2010 **2**(Suppl 1):P1.

AKos GmbH, Austr. 26, D-79585 Steinen, Germany