Chemical Non-Patent Literature Searching in E-journals and on the Internet

Maik Annies
Intellectual Property / Patent Information
Syngenta Crop Protection

Outline

- Introduction - Why searching non-patent literature is crucial
- Analysis of search and display capabilities of different E-journal full-text search sites
- Chemical structure searching in E-journals
- Chemical structure searching on the internet?

Why searching non-patent literature is crucial...

- **Scientists**: need to search non-patent literature, e.g. scientific journals, to keep up with latest publications within the scientific community

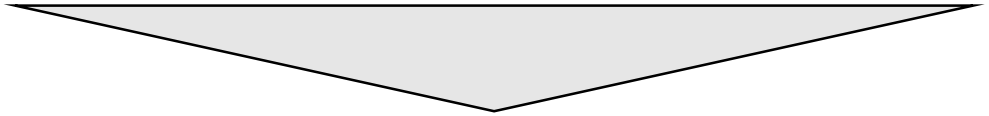
- **Patent Information Professionals**: often confronted with searches comprising analysis of non-patent literature, especially prior art searches:

  - **Novelty/patentability searches**: used to determine if a newly-conceived invention is substantially different from prior art to allow patenting
  - **Validity searches**: conducted with the intent of finding prior art to invalidate a third-party patent

Why searching non-patent literature is crucial...

Article 54 EPC

*The state of the art shall be held to comprise **everything** made available to the public by means of a written or oral description, by use, or in any other way, before the date of filing of the European patent application*



- Prior art disclosed in non-patent literature is as important as any relevant patent document
- For novelty/patentability and validity searches comparable search efforts should be undertaken for non-patent literature as for patent literature

Finding the needle in the haystack: Full-text searching of non-patent literature

Prior art is often hidden in the text, e.g. in the disclosure or in examples of publications

- Relevant prior art is often insufficiently indexed in bibliographic databases

- Novelty and validity searches should include full-text retrieval of patent AND non-patent literature!

Full-text search resources for non-patent literature prior art

Problem

- Searchable non-patent literature prior art, e.g. scientific publications in journals, meeting reports, abstracts of academic theses or websites dealing with scientific and/or technical content are scattered across different resources

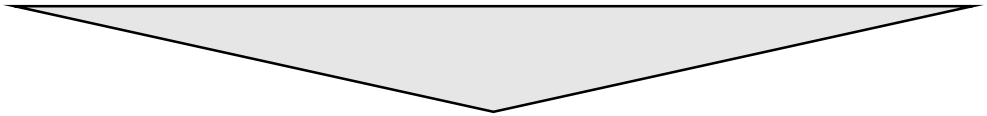
- **Online hosts**
  - Do not provide files for non-patent literature full-text searching but only for patent literature

- **Publisher E-journal search sites**
  - Often support full-text searching of E-journals

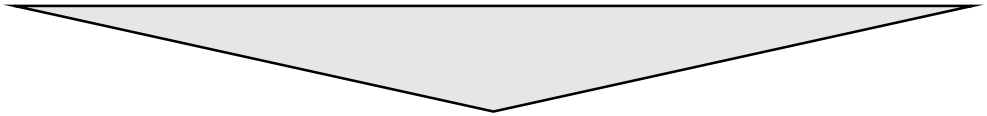
- **Internet**
  - Enormous amount of searchable scientific and/or technical information

Analysis of E-journal full-text search sites

Subscription to ~2000 E-journals at Syngenta –
mainly chemistry and (plant) biotech



Most main E-journal search sites
provide a large number of searchable journals



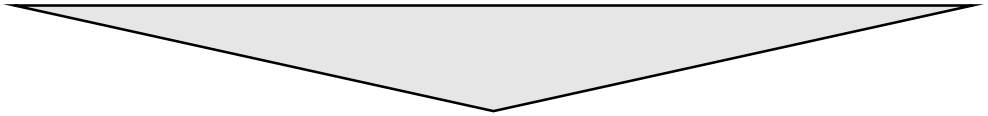
Retrieval from a large number of journals requires
searching only a limited number of search sites

Minimum requirements for E-journal full-text search sites

- Coverage of a large number of E-journals, ideally from as many different publishers as possible, dealing with technologies of relevance
- Advanced search options - at least Boolean logic or wildcards should be supported
- Advanced display features – at least searched keywords should be highlighted within the context of the full-text

Detailed evaluation of E-journal full-text search sites

Selection of E-journal full-text search sites based on minimum requirements – coverage, search options and display features

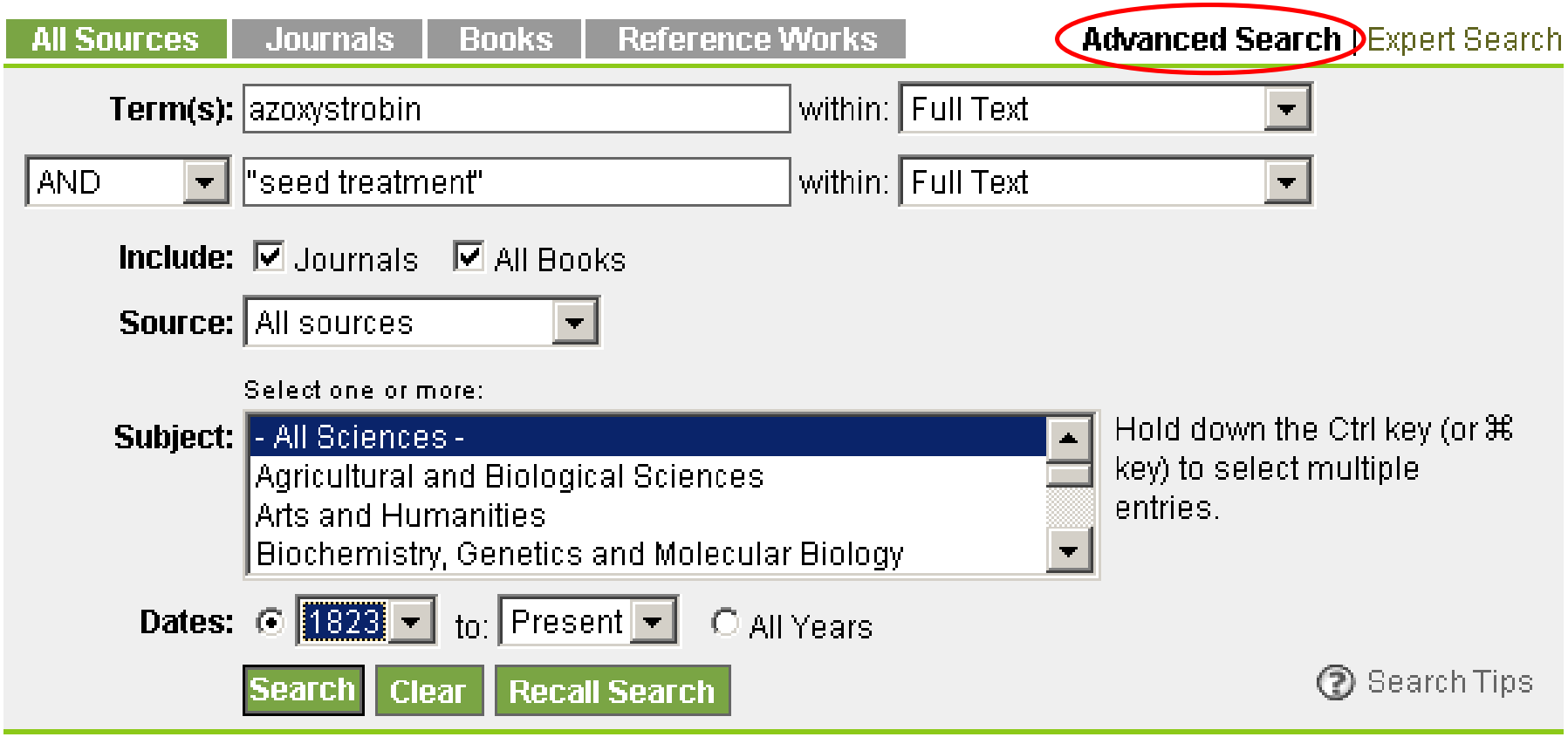


Only 5 E-journal full-text search sites fulfilled minimum requirements

E-journal full-text search sites fulfilling minimum requirements

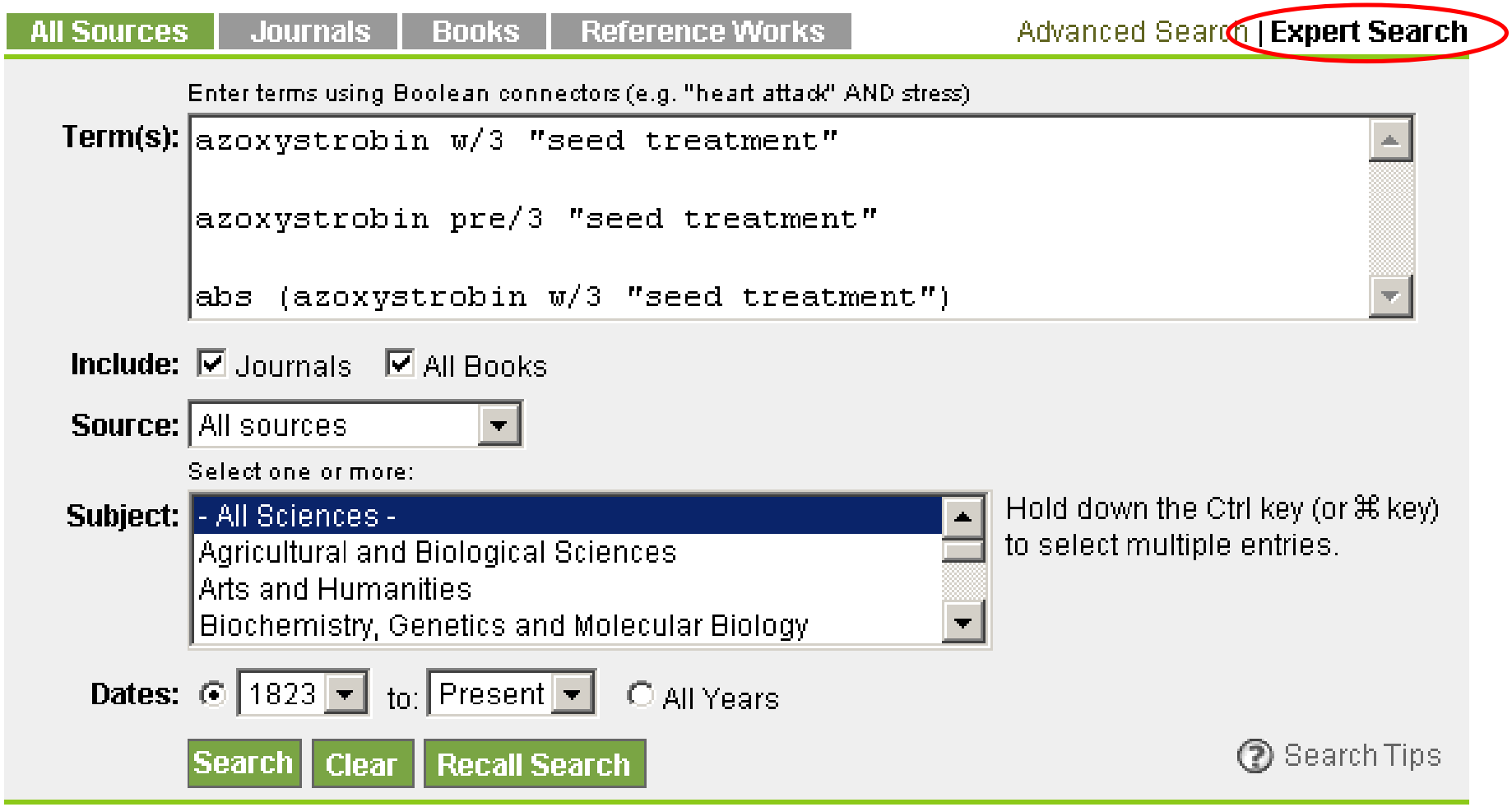
	Number of Journals
ACS Legacy Archives	37
Annual Reviews	37
Blackwell Synergy	~ 850
Highwire Press	~ 1100
Science Direct (Elsevier)	~ 2500

Search options – example 1



Example from www.sciencedirect.com (Elsevier)

Search options – example 2



Example from www.sciencedirect.com (Elsevier)

Search options - overview

	Boolean Logic	Phrase Searching	Stemming	Wildcards	Proximity Operators
ACS Legacy Archives	√	√	√		
Annual Reviews	√	√	√	√	
Blackwell Synergy	√	√	√		
HighWire Press	√	√	√	√	
Science Direct (Elsevier)	√	√		√	√

Display features – lack of multicoloured keyword highlighting

Seed treatment (ST) with agrochemicals has been an effective method to control a variety of insect pests and diseases (*1*). Systemic materials, such as **tebuconazole**, triticonazole, fludioxonil, silthiofam, imidacloprid, **thiamethoxam**, and **fipronil**, provide very good broad spectrum activity and excellent control of diseases and insect pests, particularly in early crop growth stages. These systemic pesticides also have less hazardous toxicological and ecotoxicological profiles than commercially applied broadcast or in-furrow options. The major crops that benefit from the use of ST are cereals, maize, cotton, potatoes, oilseed rape, and sugar beet. **Seed treatment** actually reduces the amount of active ingredient applied to the environment because **treatment** is restricted to the target crop. Although several insecticides and fungicides available for ST are systemic in nature, they typically can provide protection from diseases and pests only to the seed and young seedlings.

Example from pubs.acs.org (ACS Legacy Archives)

J. Agric. Food Chem., 52 (15), 4814-4820, 2004

Display features – do not always trust in automatic stemming

The effects of fungicide **seed** treatments on plant populations and subsequent incidence of tomato spotted wilt have been examined. Brenneman & Walcott (21) reported that use of **seed treatment** combinations of carboxin, captan, and **pentachloronitrobenzene** (PCNB) increased plant stands by almost 60% compared to the nontreated control, which corresponded with an approximately 60% reduction in final incidence of spotted wilt. That combination of fungicides is a standard **seed treatment** used on most peanuts planted in the United States. Rideout and associates (115) reported no improvement in stand and no significant reduction in incidence of spotted wilt with in-furrow applications of the fungicide azoxystrobin in addition to the standard carboxin, captan, and PCNB fungicide **seed treatment**.

Example from arjournals.annualreviews.org (Annual Reviews)

Annu. Rev. Phytopathol., 41, 53-75, 2003

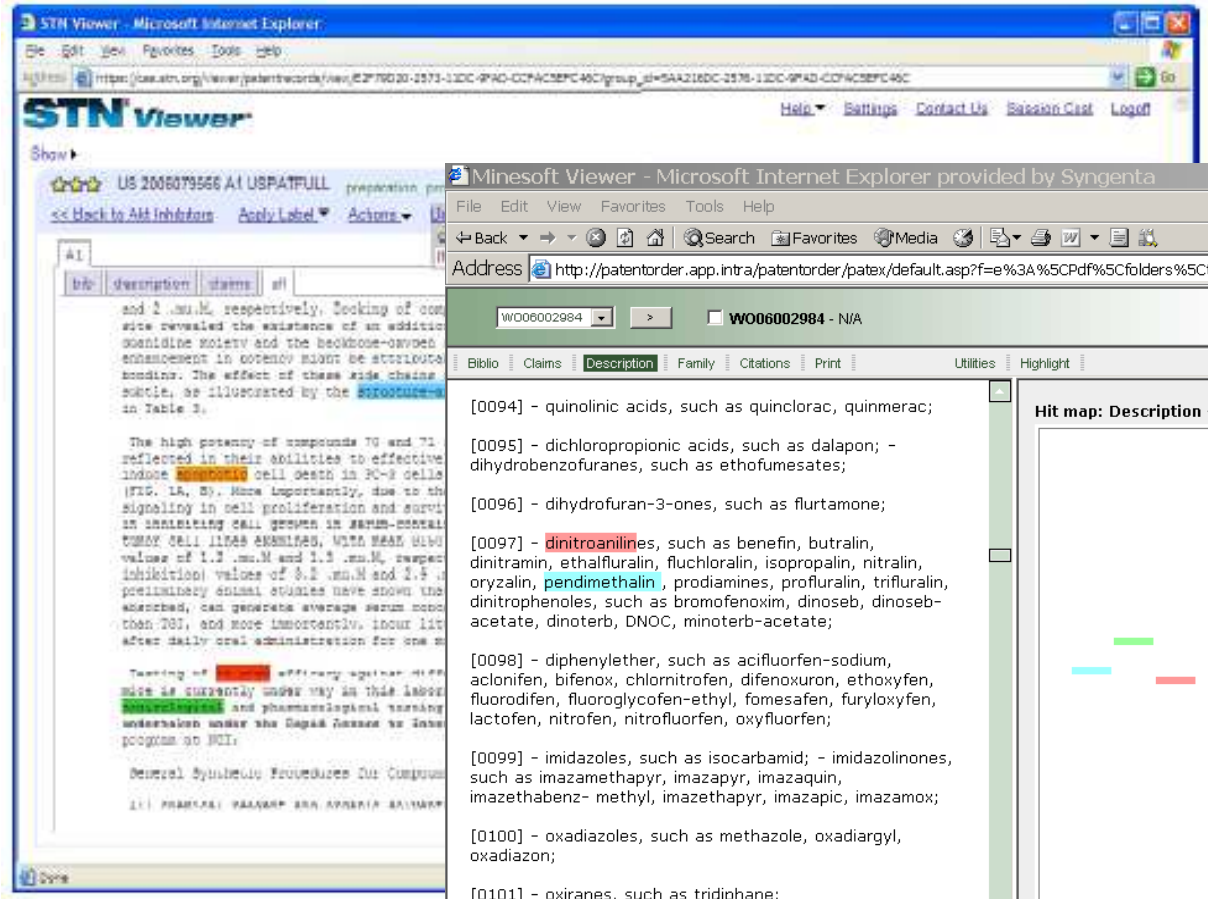
Display features – be careful with molecular formulas

For example, **ethanol** (C_2H_5OH) and formic acid (HCOOH) are both detected at an M + 1 value of 47 amu. The natural relative isotopic ^{13}C abundance of protonated **ethanol** is 2.3% (with two C atoms present), while for protonated formic acid, it is only 1.2% (one C atom present). In the *Candida tropicalis* experiments, a relative isotopic ^{13}C abundance of 2.4% was observed (Fig. 2c), indicating that **ethanol** was produced. It should be pointed out that isotopic pattern analysis cannot be used to discriminate between isomeric compounds, i.e., compounds with the same molecular weight and the same sum formula. Dimethyl ether (CH_3OCH_3) is isomeric to **ethanol** and cannot be distinguished from the alcohol by PTR-MS. The microbial production of this ether in large quantities, however, can be excluded with high confidence.

Example from highwire.stanford.edu (Highwire Press)

Appl. Environ. Microbiol. 74 (7), 2179-2186, 2008

The ideal solution – yet only available for patents



STN Viewer

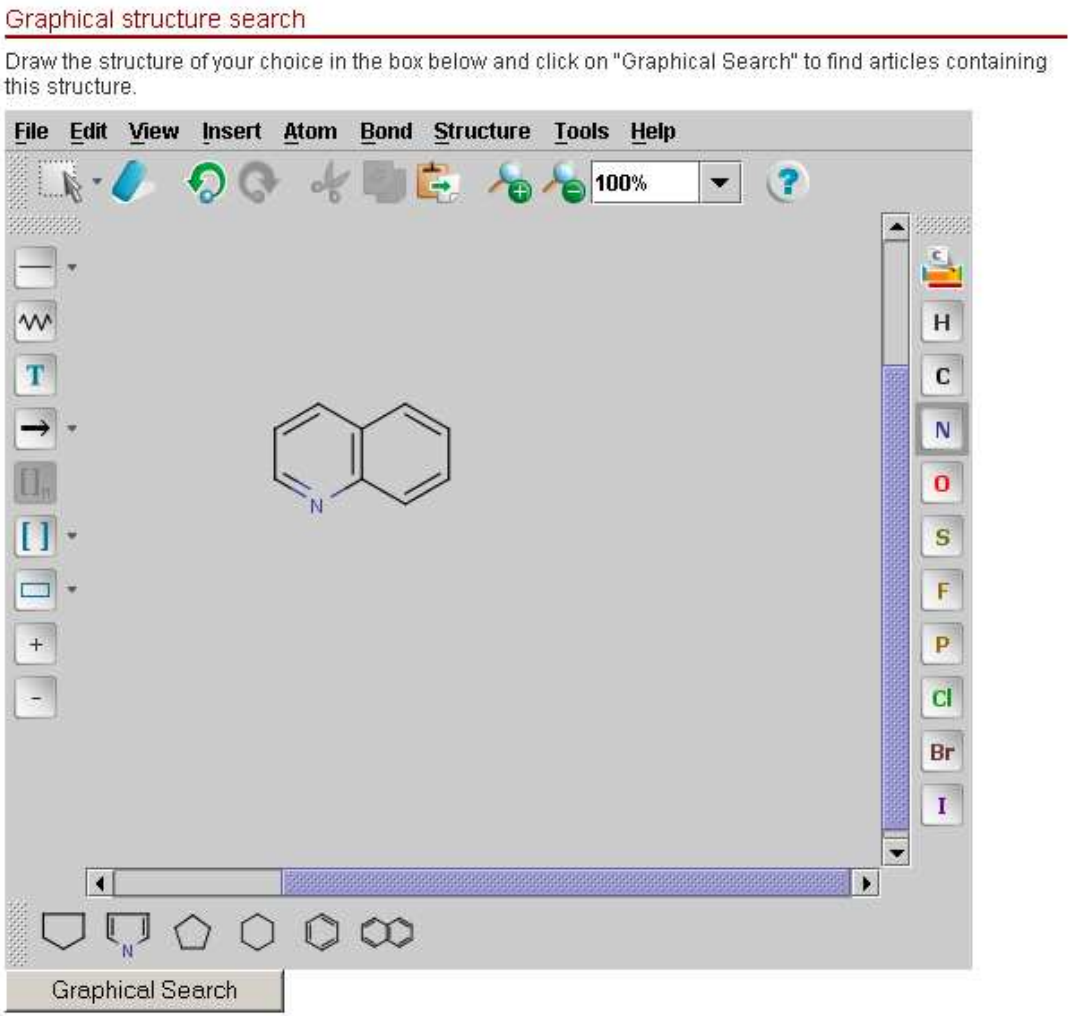
Minesoft PatentViewer

syngenta

Display and other features - overview

	Multi-coloured Keywords	Save Function	Search Alerts	Open Access
ACS Legacy Archives				
Annual Reviews		√	√	
Blackwell Synergy		√	√	√
HighWire Press		√	√	√
Science Direct (Elsevier)		√	√	√

Chemical structure searching on E-journal search sites



RSC Publishing

PROUS SCIENCE

Example from www.rsc.org (Royal Society of Chemistry)

syngenta

Chemical structure searching on E-journal search sites

- Only very few publishers (RSC Journals and Prous Science) provide chemical structure searching in E-journals
- Both exact searching and substructure searching in E-journals are provided by both publishers
- Currently only a very limited number of E-journals is at all structure-searchable

Chemical structure searching on the internet? – Free web-based search resources

- Some web-based structure search resources are available on the internet providing access to substance-related information
- No comprehensive retrieval of bibliographic data or link to chemical structures on the web

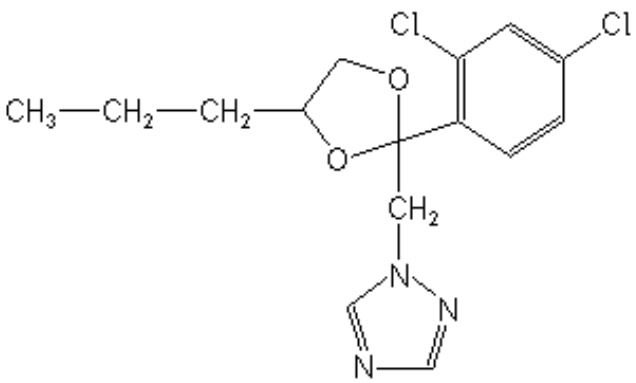
Chemical structure searching on the internet? – The InChI-approach

Solution

- Information from chemical structures needs to be transformed into an alpha-numerical representation indexable by web search engines and thus searchable, which **requires to agree on a common standard**

- Most promising candidate:
  InChI (IUPAC International Chemical Identifier):
  - Developed by IUPAC and NIST during 2000-2005
  - Contains chemical information about e.g. bond connectivity, tautomerism, isotopes, stereochemistry, and electronic charge
  - Non-proprietary, unambiguous identifier that can be computed from structural information
  - Already supported by some databases and scientific journals

Chemical structure searching on the internet? – The InChI-approach



InChI=1/C15H17Cl2N3O2/c1-2-3-12-7-21-15(22-12,8-20-10-18-9-19-20)13-5-4-11(16)6-14(13)17/h4-6,9-10,12H,2-3,7-8H2,1H3

STJLVHWMYQXCPB-UHFFFAOYAJ (**InChIKey**)

- Principally InChI/InChIKey for different chemical compounds are precisely indexed by major web search engines (Cole et al., 2005)
- However, retrieval is not satisfactory, since InChI are not yet sufficiently used by authors/publishers of journals and web content!

Chemical structure searching on the internet? – Chemical data mining

Problem

- Automatic transformation of (non)-standardized nomenclature or non-computer-readable formats, e.g. images of chemical structures, into searchable computer-readable formats

Solution

- Automatic chemical name to structure conversion or graphical structure recognition tools are already available or under development

- Storage of these computer-readable formats in structure databases and linking to original websites would allow (sub)structure searching of chemical content on the web

Thank you

for your attention!