Europäisches Patentamt
European Patent Office
Office européen des brevets

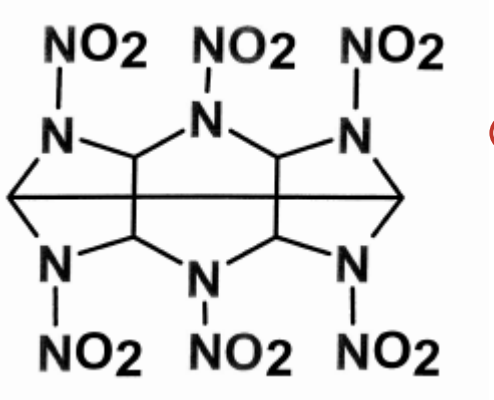
Software Assisted Mental Representations of Chemistry Information Retrieval Tasks

Dr. Stefan Härtinger, European Patent Office, Munich

ICIC 2012 Conference, 14-17 October 2012, Berlin

Patent examiner

DISCLAIMER: The views or opinions expressed by the speaker are solely his own and do not necessarily represent the views or opinions of the European Patent Office.



What comes first to your mind when looking at this chemical formula?

How would you search for it?

Mental Representation

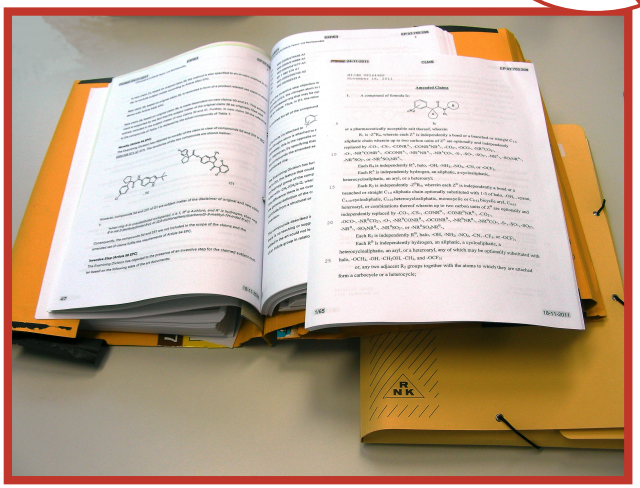
Motivation 1:

Mental Representation of Chemistry & Info. Retrieval Task

Chemistry Information

Domain & Search Knowledge

Software



C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to c
X	GB 1 203 328 A (TREVOR JOHN FRANKLIN) 26 August 1970 (1970-08-26) example 1 ---	1,2,6-
X	WO 94 12184 A (SYNTEX INC) 9 June 1994 (1994-06-09) examples 1,2 ---	1,10
X	US 5 688 529 A (HEGDE SAYEE GOJANAN ET AL) 18 November 1997 (1997-11-18) abstract ---	1,7,10
Y	WO 97 38689 A (HAEBERLIN BARBARA ;CIBA GEIGY AG (CH); MAK CHING PONG (CH); MEINZE) 23 October 1997 (1997-10-23) cited in the application example 1 --- -/--	1-10

Patent application

Patent examiner

Search report

Patentability Search

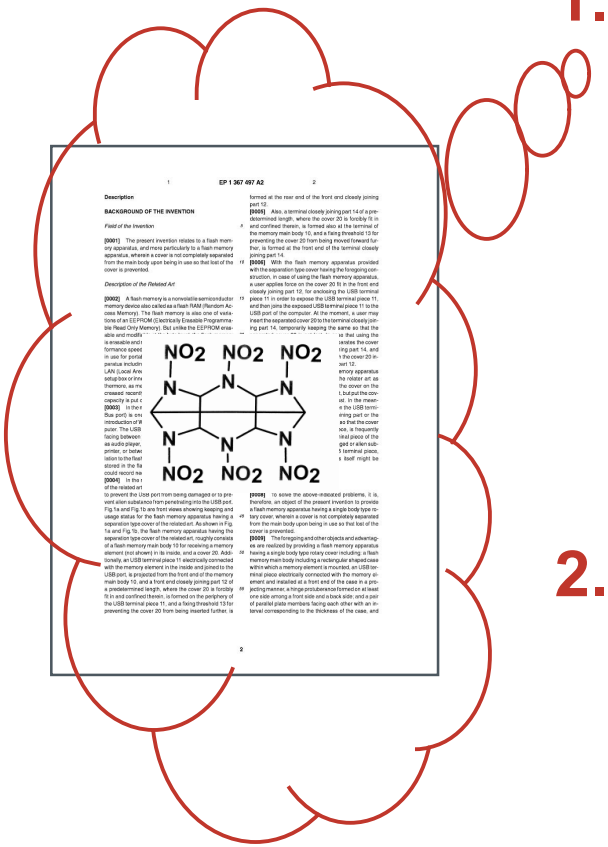
Motivation 2:

„…Search phase will be ***automated*** with the purpose to automate as far as possible the search process and eliminate all ***non-value added steps for examiners*** at the beginning of the search workflow. Successive steps in the project will deliver ever more relevant documents before the examiner opens the dossier…“.

Source: IT Roadmap of the EPO

- User = Patent Examiner
- Information Retrieval Task = Patentability Search
- User-centric, User Observation, User Representative in Teams

Session Outline:



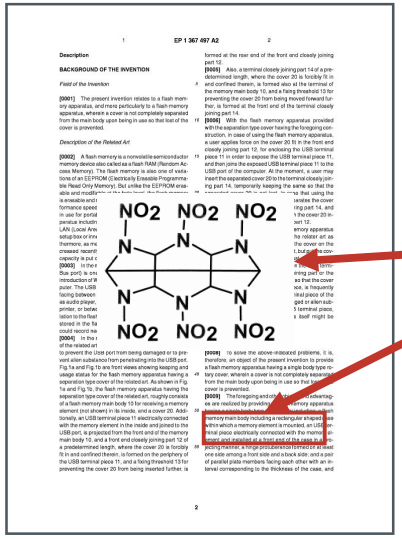
1. Mental Representations

- How do they evolve from chemical information?
- How do they affect the information retrieval?

2. Software

- How to support the human way of searching?

Mental Representations of Chemical Compounds



EP 1 367 497 A2

Description

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention relates to a flash memory apparatus, and more particularly to a flash memory apparatus, wherein a cover is not completely separated from the main body upon being in use so that lost of the cover is prevented.

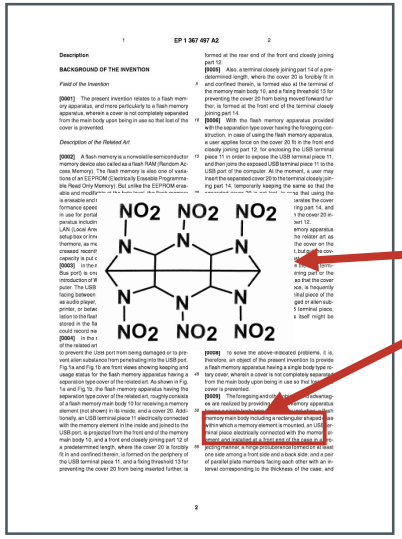
Description of the Related Art

[0002] A flash memory is a nonvolatile semiconductor memory device also called as a flash RAM (Random Access Memory). The flash memory is also one of variations of an EEPROM (Electrically Erasable Programmable Read Only Memory). But, unlike the EEPROM erasable and modifiable

Chemical information

How do mental representations evolve from information?

Mental Representations of Chemical Compounds



Chemical information

Recognise

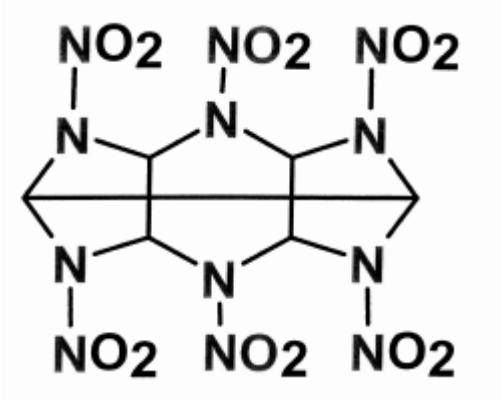
Understand

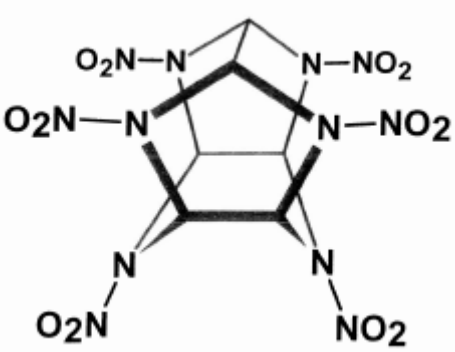
Formulate search task

Search = technical procedure (same procedure always **same** results)
Searchers = human beings (same procedure **equivalent** results)

(Adopted from van de Kuilen, „When is a search or a searcher good enough?“ Pat. Inf. Conf., 2008, Stockholm)

Problems in understanding chemistry?



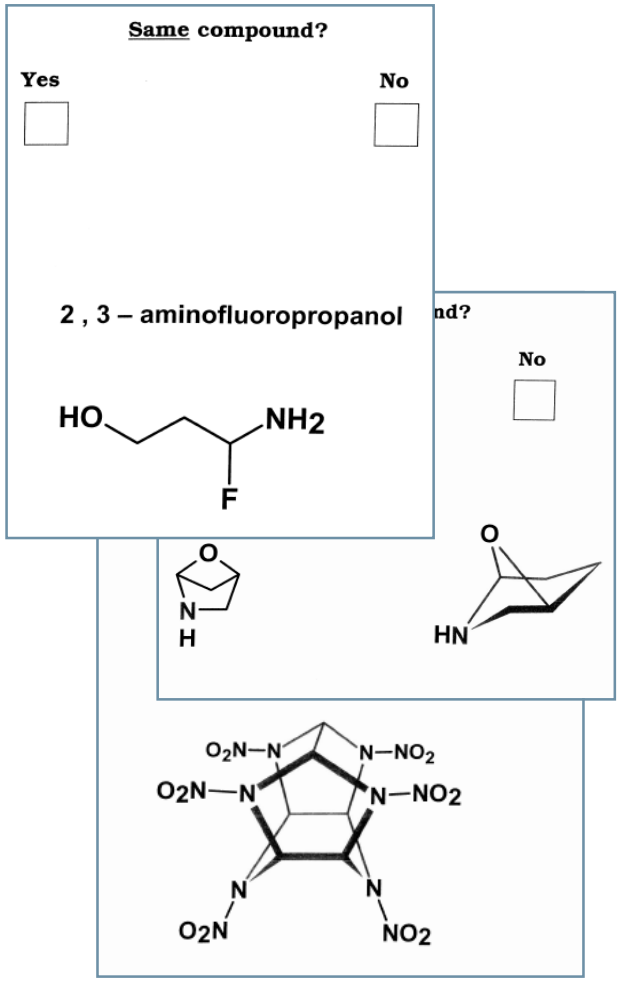


Same compound?

Yes ☐ **No** ☐

Cognitive overload / Error-prone / Time-consuming

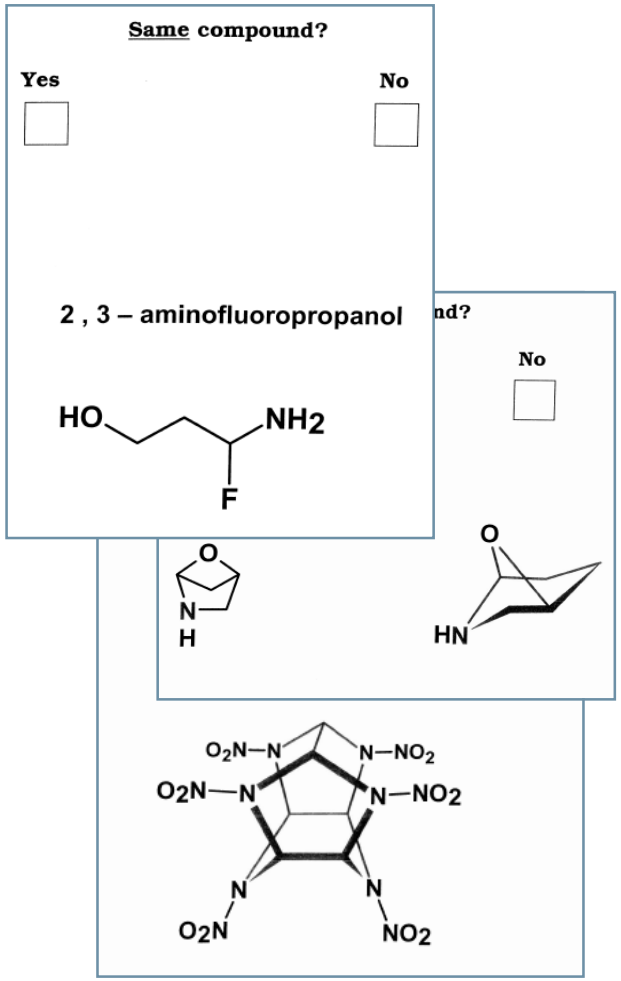
Everyone struggles



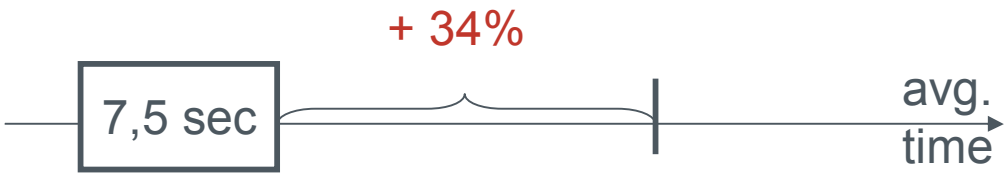
Name *vs.* Structure

2D *vs.* 3D

Everyone struggles …but in a different manner!



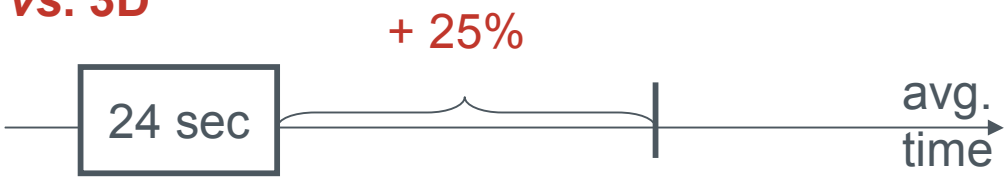
Name *vs.* Structure



Pharmacology
50% correct

Organic Chemistry
100% correct

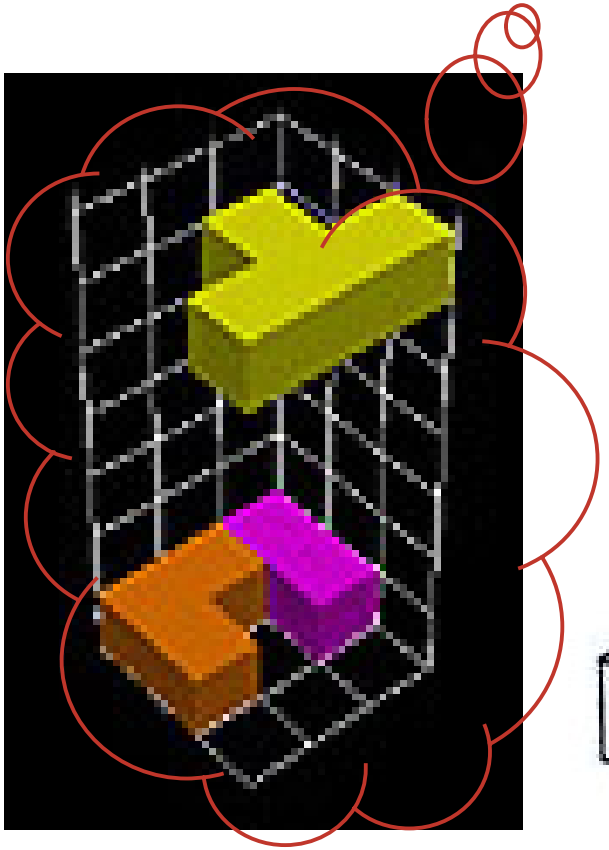
2D *vs.* 3D

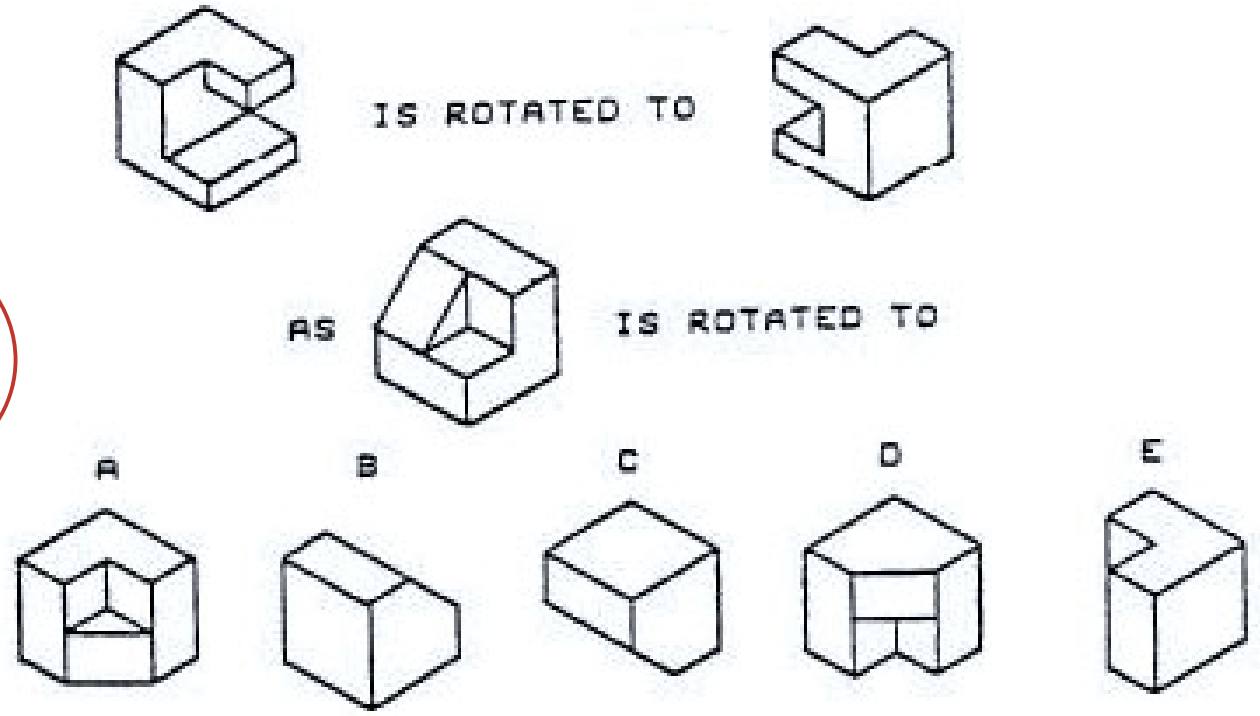


Organic Chemistry
100% correct

Pharmacology
100% correct

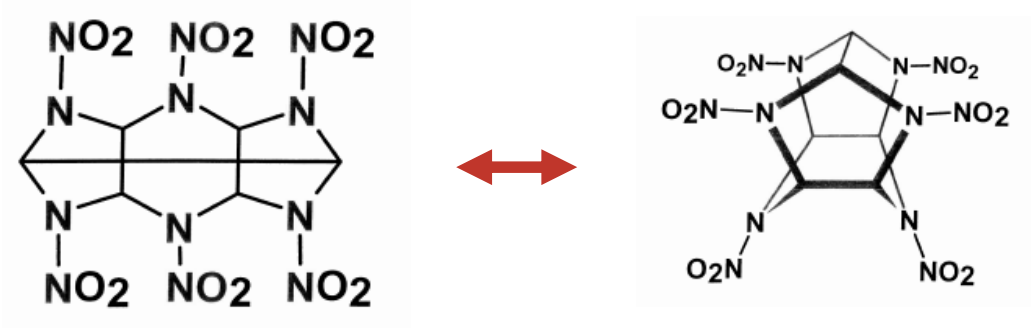
“Tetris for chemists”: 3D spatio-visual ability



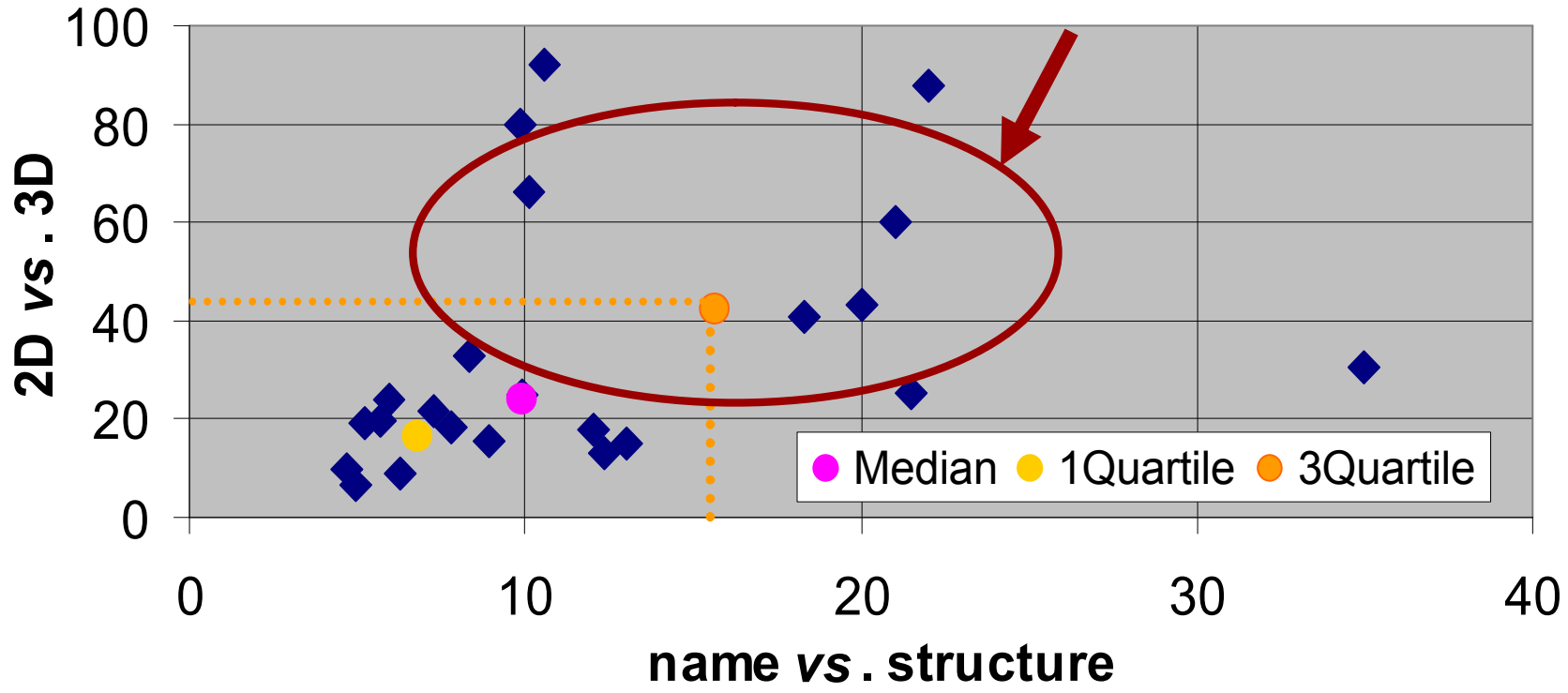


„…The **correlation** between **spatial ability** and students performance on **chemistry** exams was **significant** for questions that required **problem-solving skills**…“.

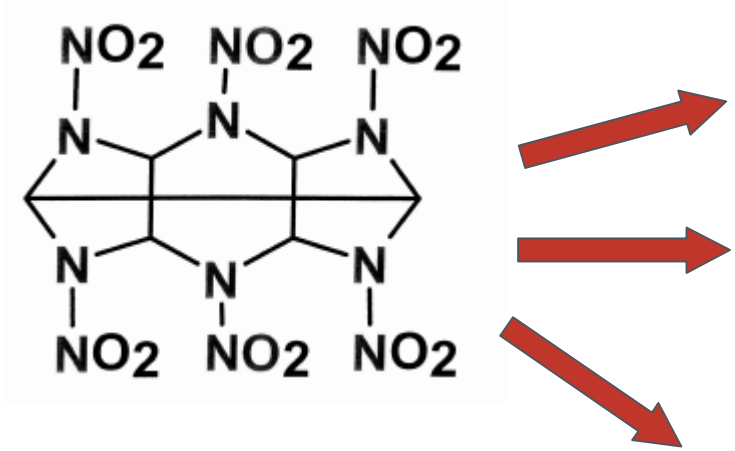
Bodener et.al. „The Purdue visualization of rotations test“, Chemical Educator, 2 (1997) 1-16

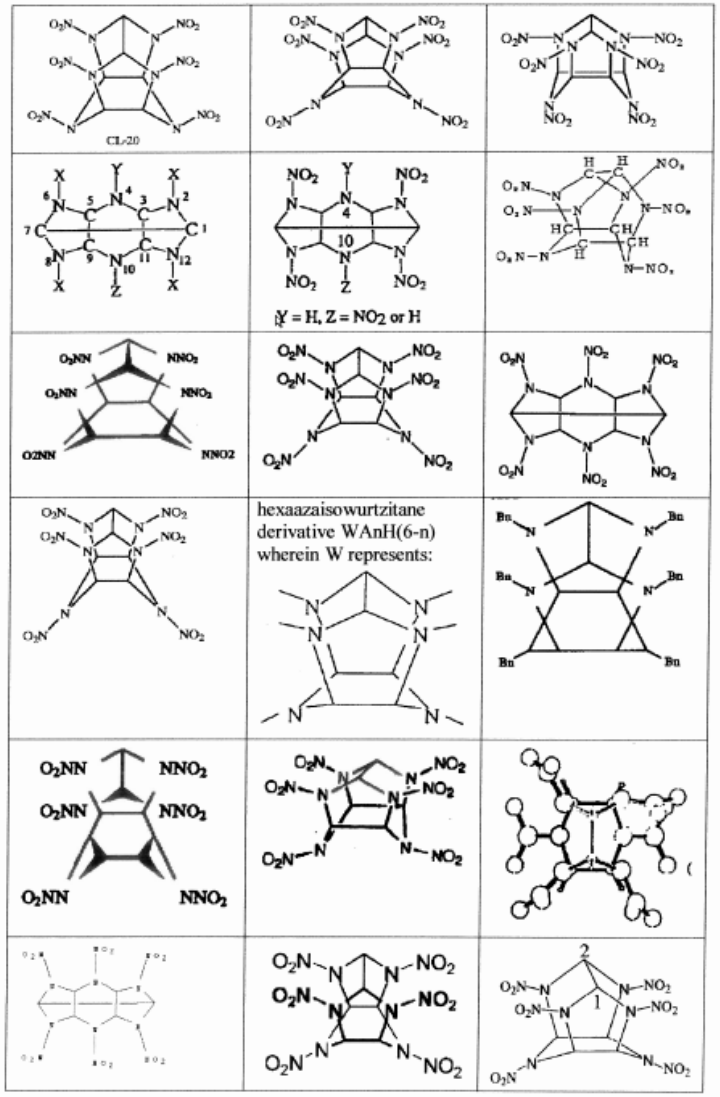


Population of „low“ spatio-visual ability



High cognitive workload



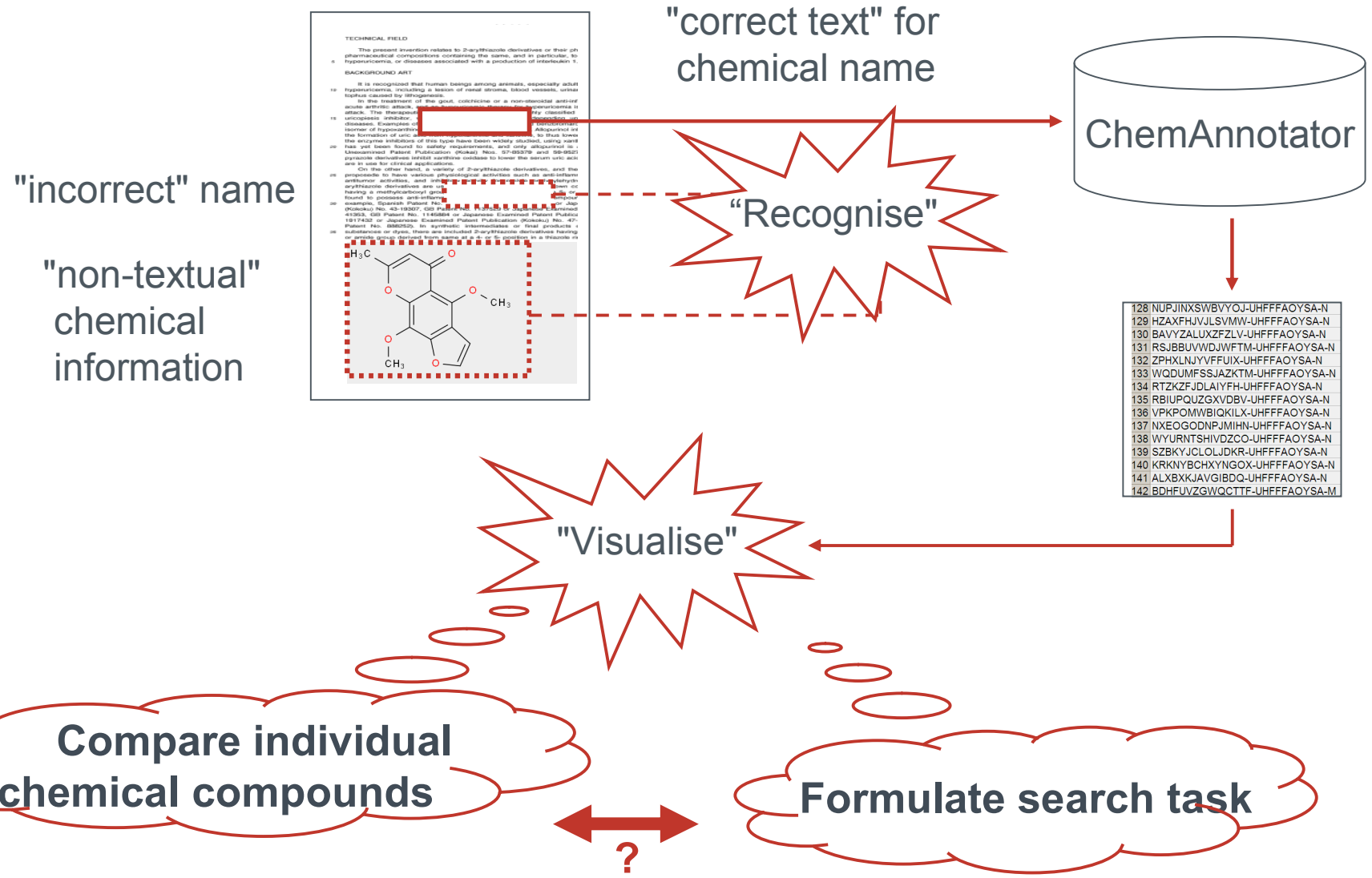


Repeated comparision with mental model
≈ 200 documents à ≈ 44 pages
Efficiency problem

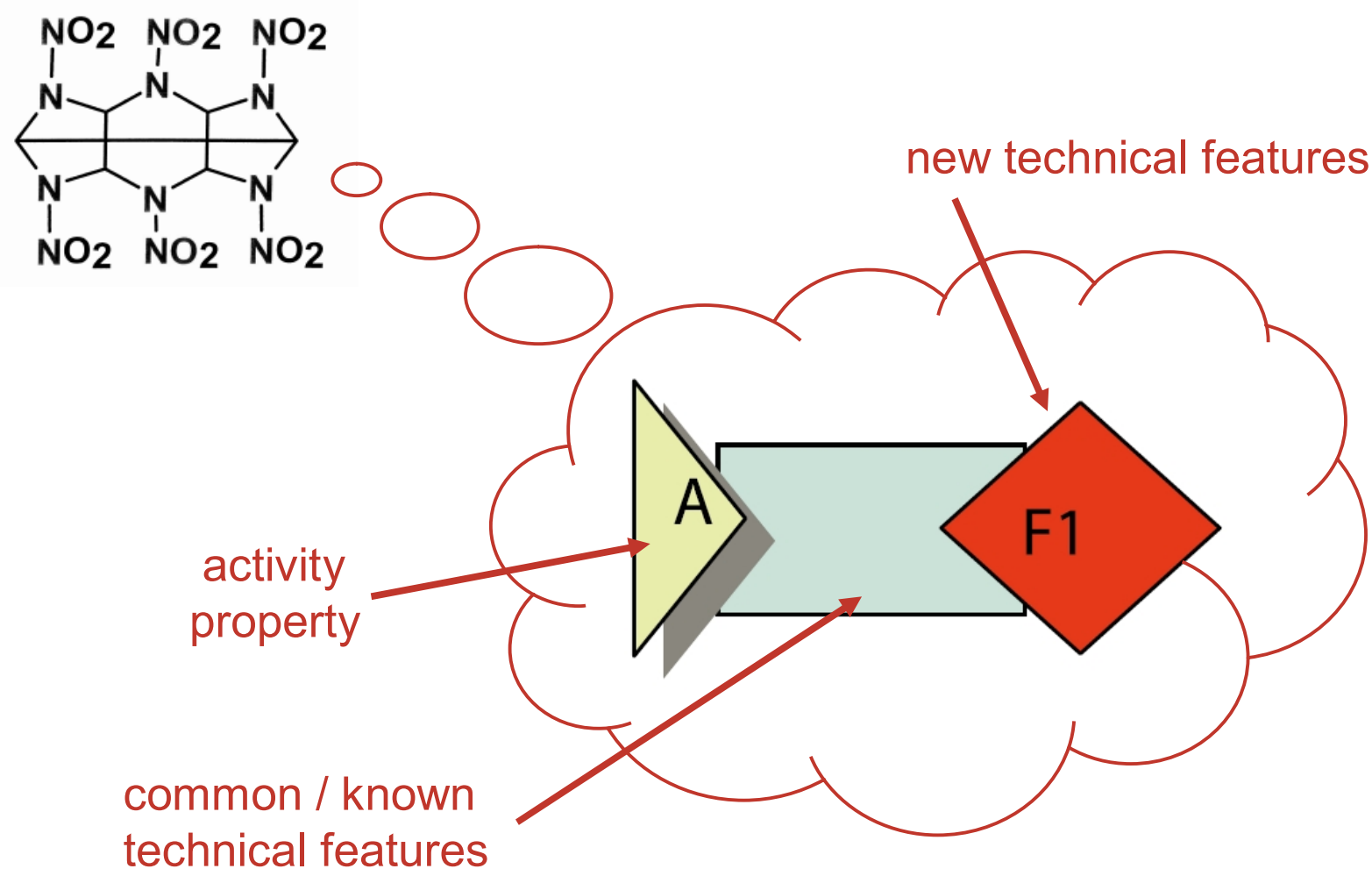
Not everyone equally suited
Quality challenge

- **Software assistance needed**

EPO Chemical annotation tool

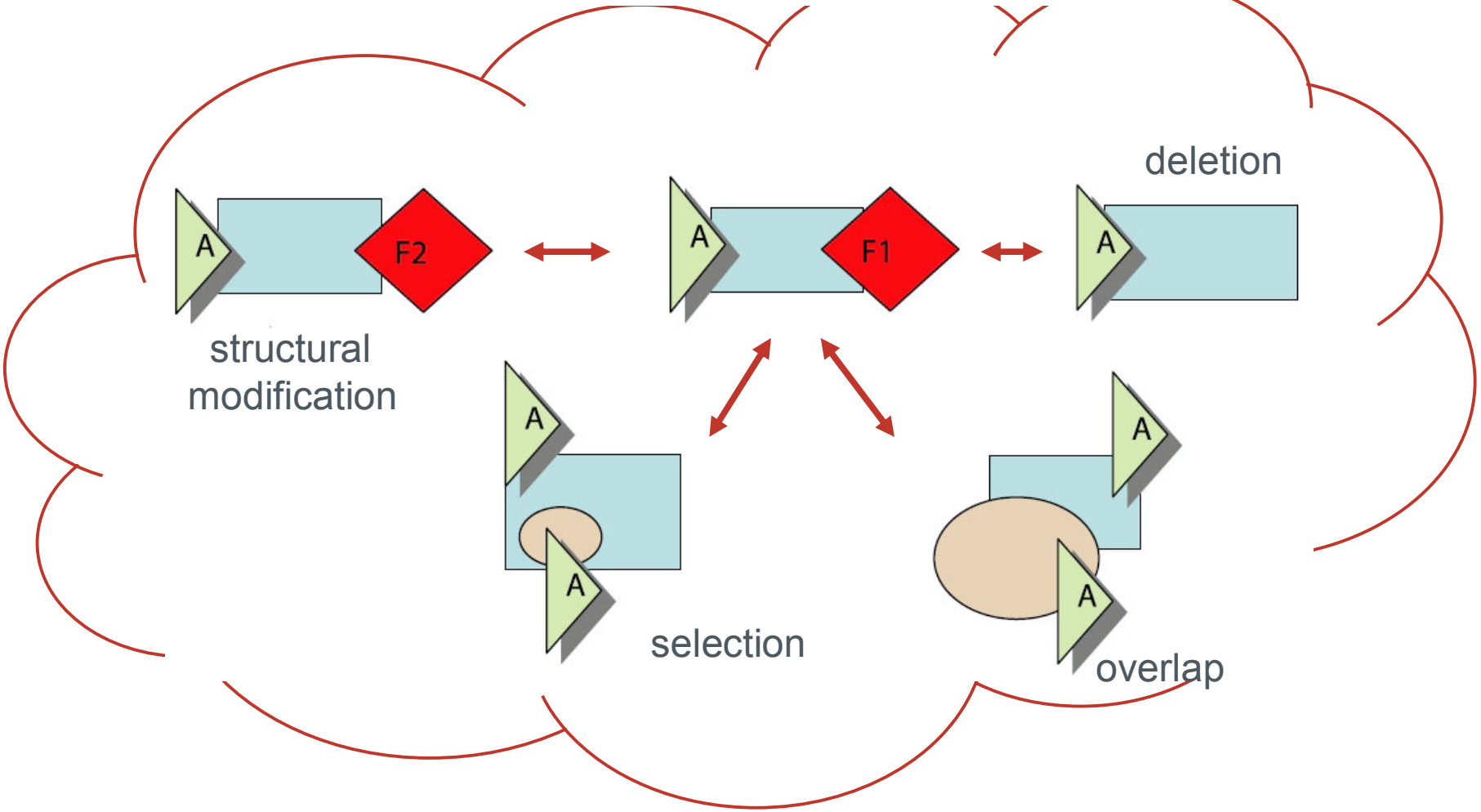


Differences in the chemical space – legal constraints



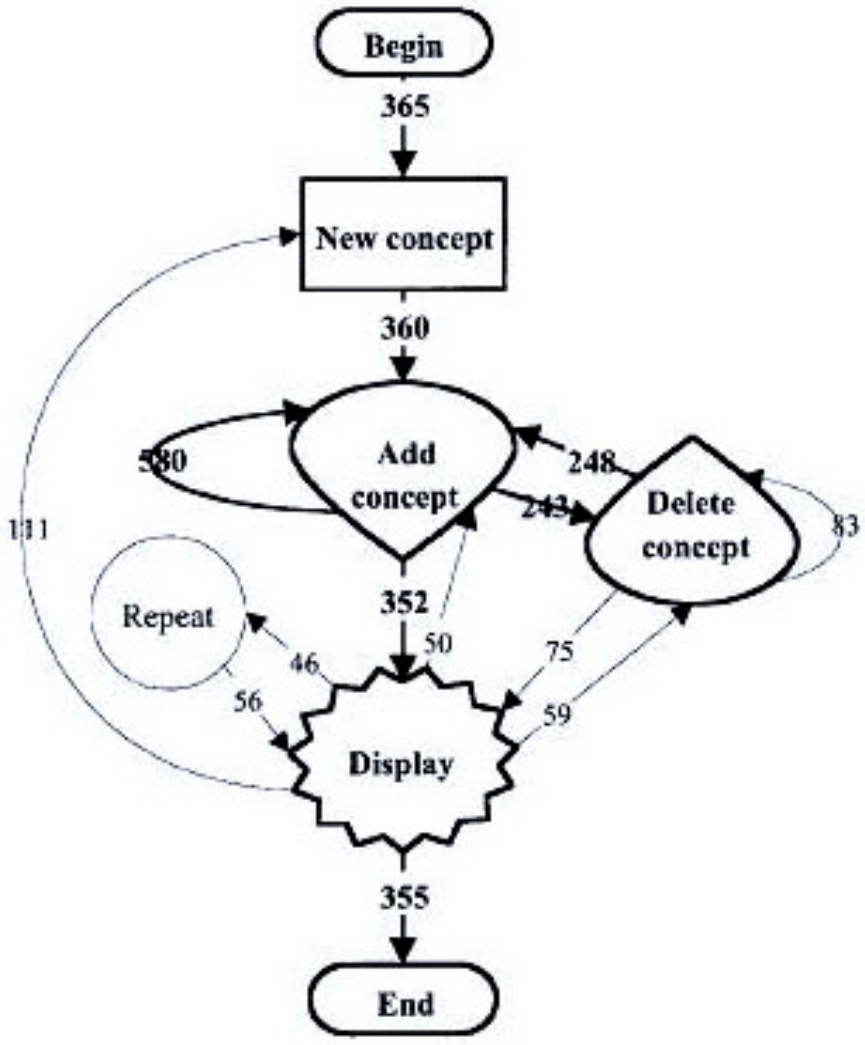
Stellmach J.A. „A graphical representation of the problem-solution approach to the assessment of inventive step“ World Patent Info., 31 (2009) 4-10

Formulation of the information retrieval task



Repeated comparisons & refinements & adjustments of the mental representations as the search progresses and “closer” hits are retrieved

Dynamic changes of mental representations



- Apparent implicitly through formulation of search statements

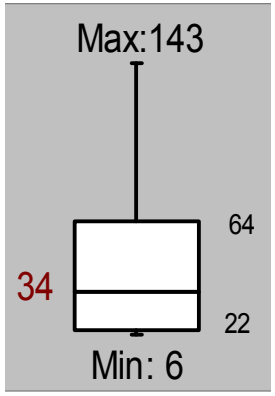
  - key words
  - structure queries
  - classification symbols

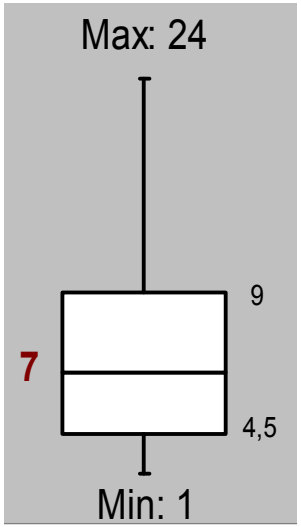
- Logfile („transaction log“) automatically created

- Non-intrusive analysis of searchers´ activity

Wildemuth B.M. „The Effects of Domain Knowledge on Search Tactic Formlation“, J. Am. Soc. Info. Sci. Tech., 55 (2004) 246-258

Dynamic changes of mental representations





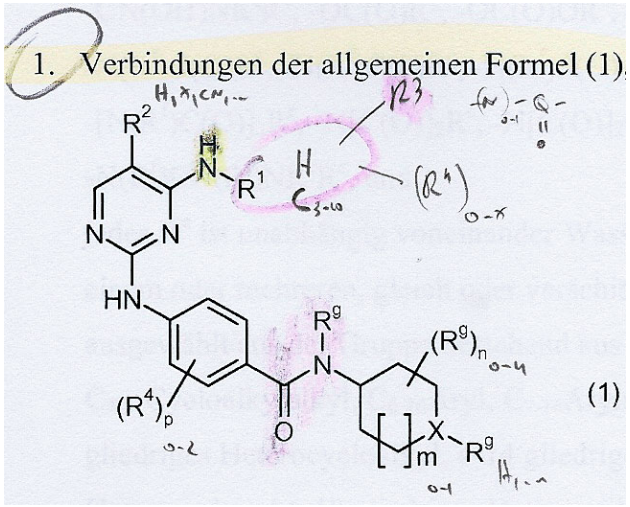
Patentability searches with organic chemistry and/or pharmacology content

- **34 different search concepts**
  (median of skewed distribution)

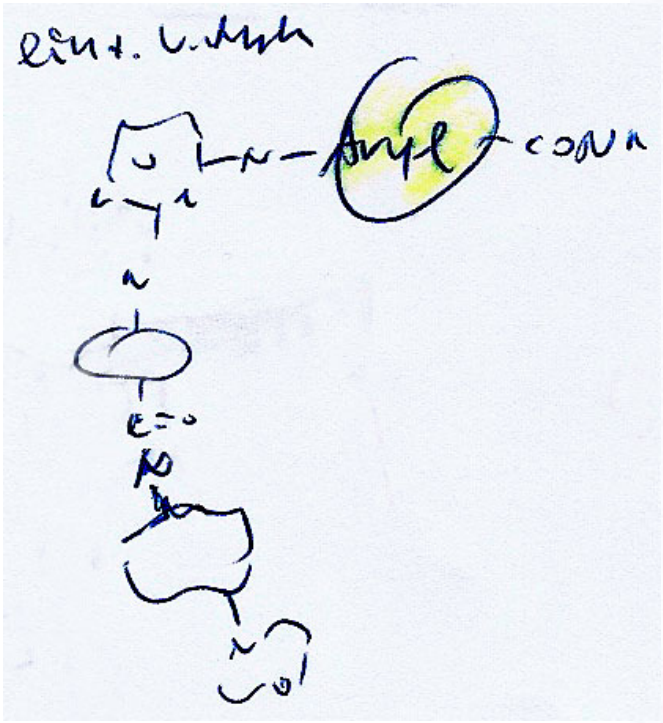
- **7 search strategy repeats/modifications in different databases/clusters**
  (median of skewed distribution)

Search Progress

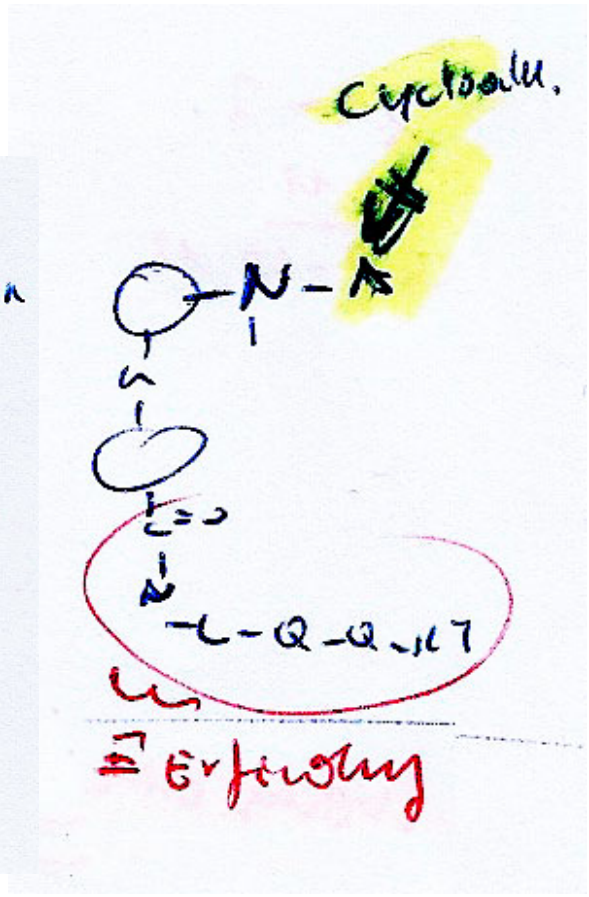
Change in mental Representations



Analysis of invention



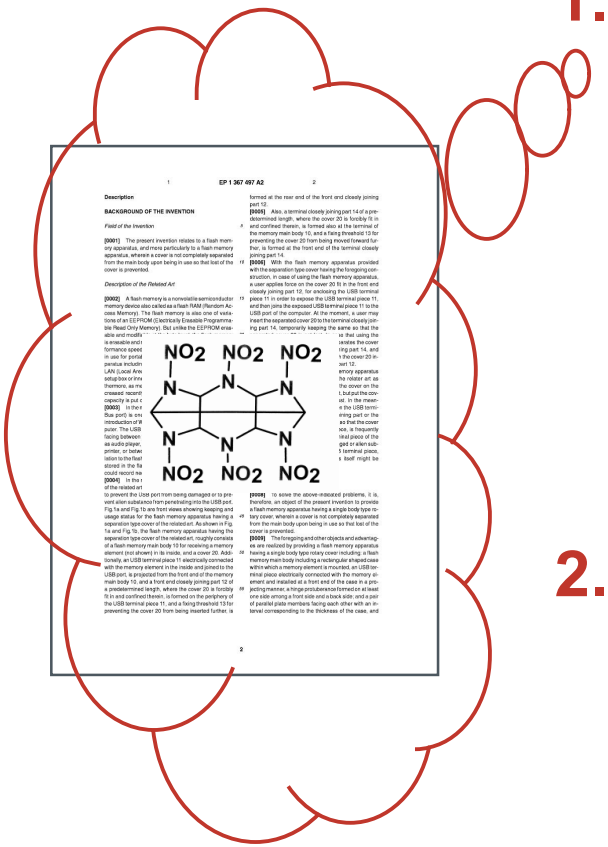
Prior art found during initial pre-search



Prior art found at final search stage

Search Progress

Session Outline:



1. Mental Representations

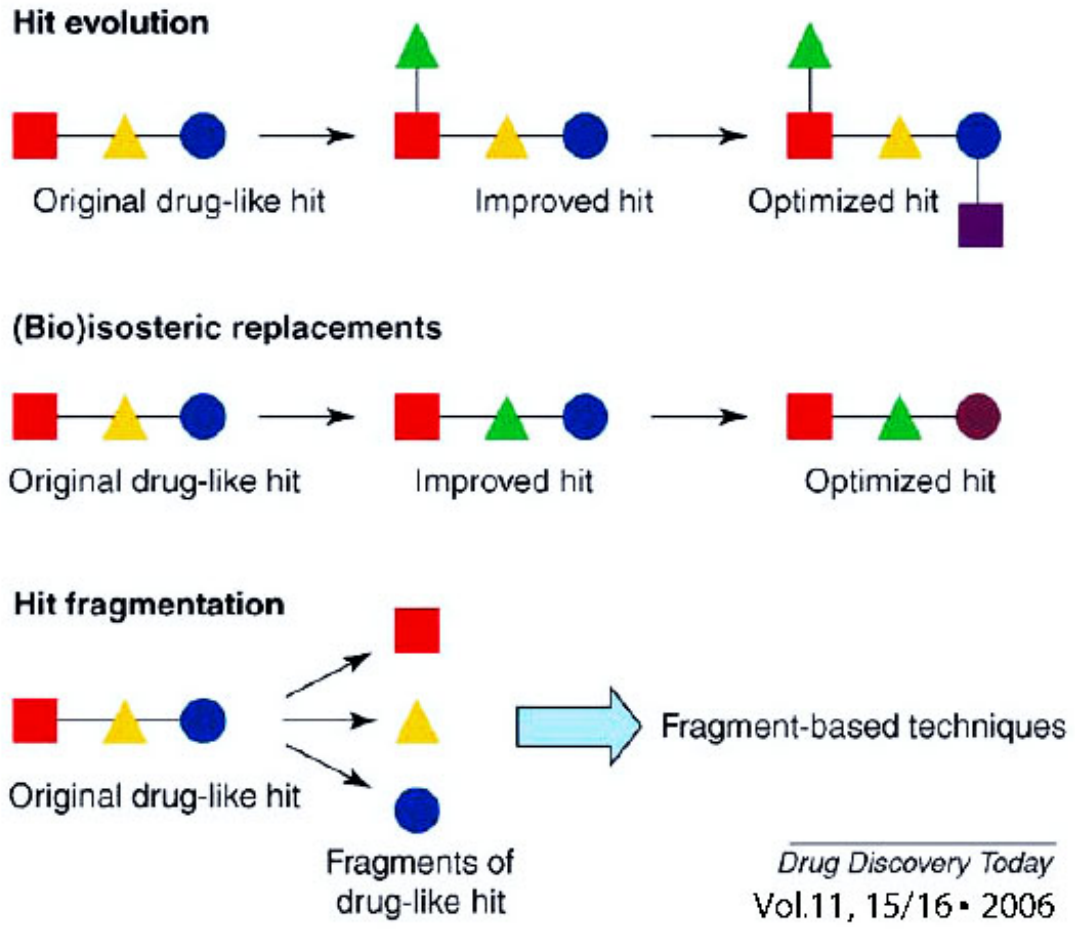
- How do they evolve from chemical information?

- How do they affect the information retrieval?

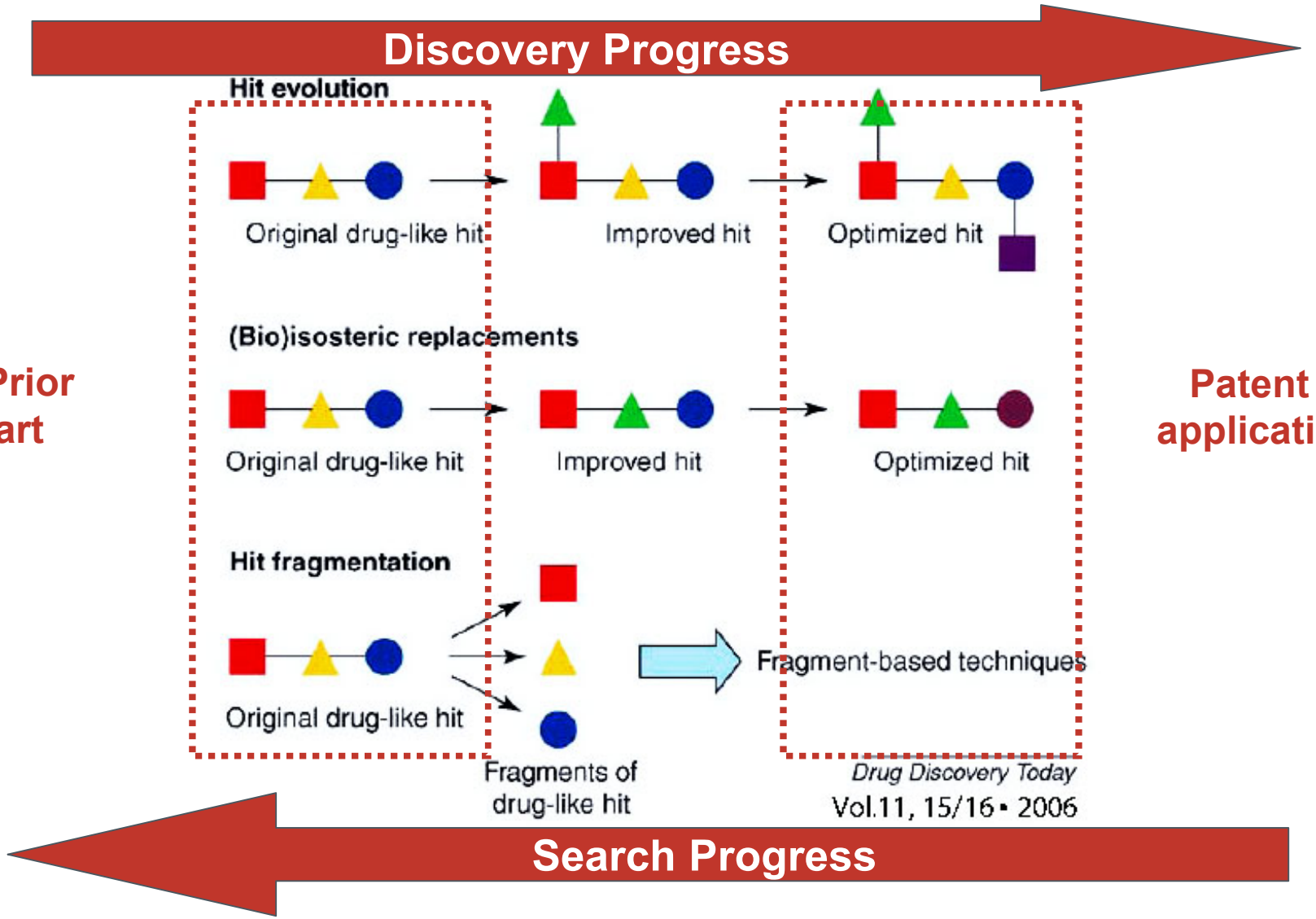
2. Software

- How to support the human way of searching?

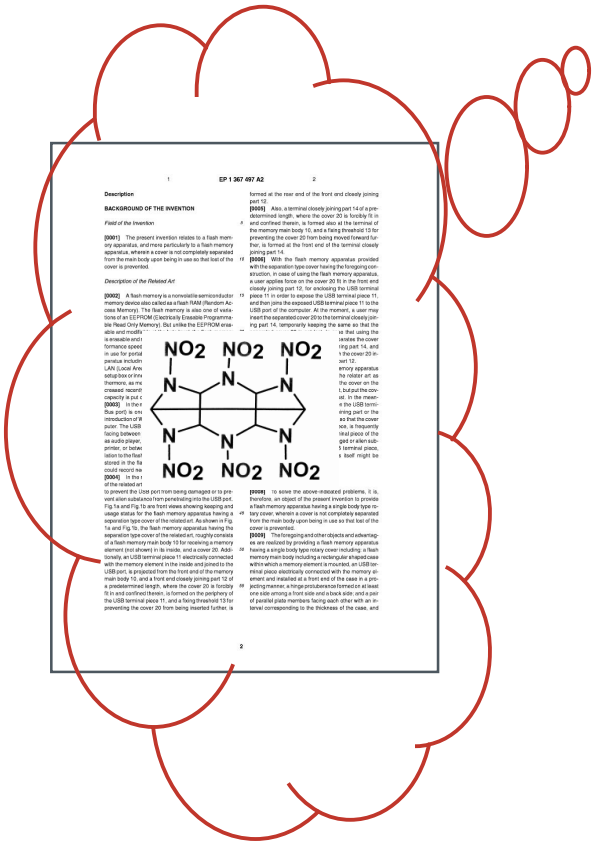
Discovery of new small molecule entities and drugs



Discovery of new small molecule entities and drugs



Software

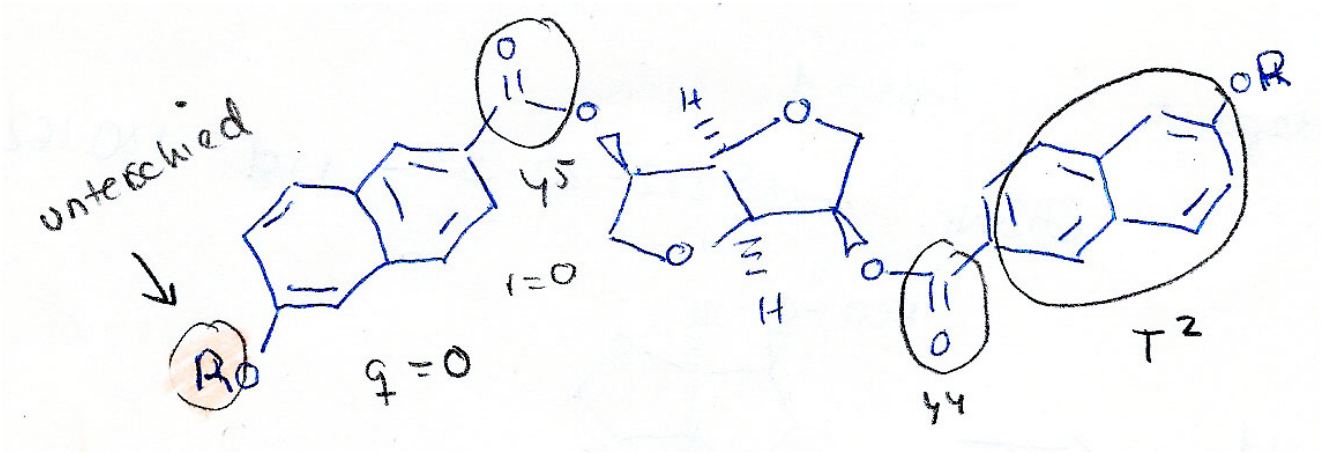


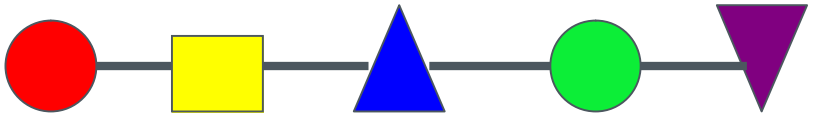
…support the human in building and using mental representations

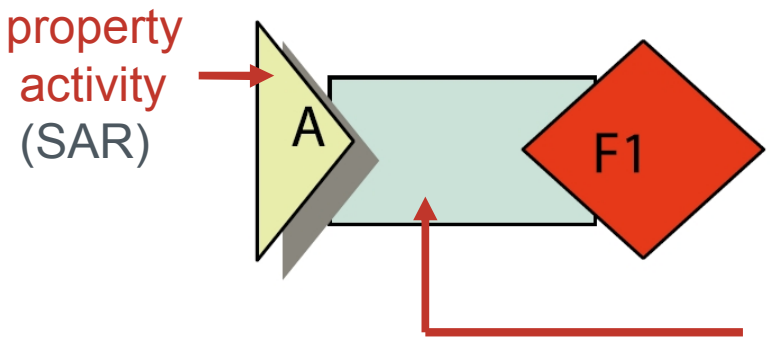
- Highlighting
- Searching
- Grouping

- Monitoring progress
- Guidance on strategy
- Documenting results
- ….

Tools to highlight distinguishing features



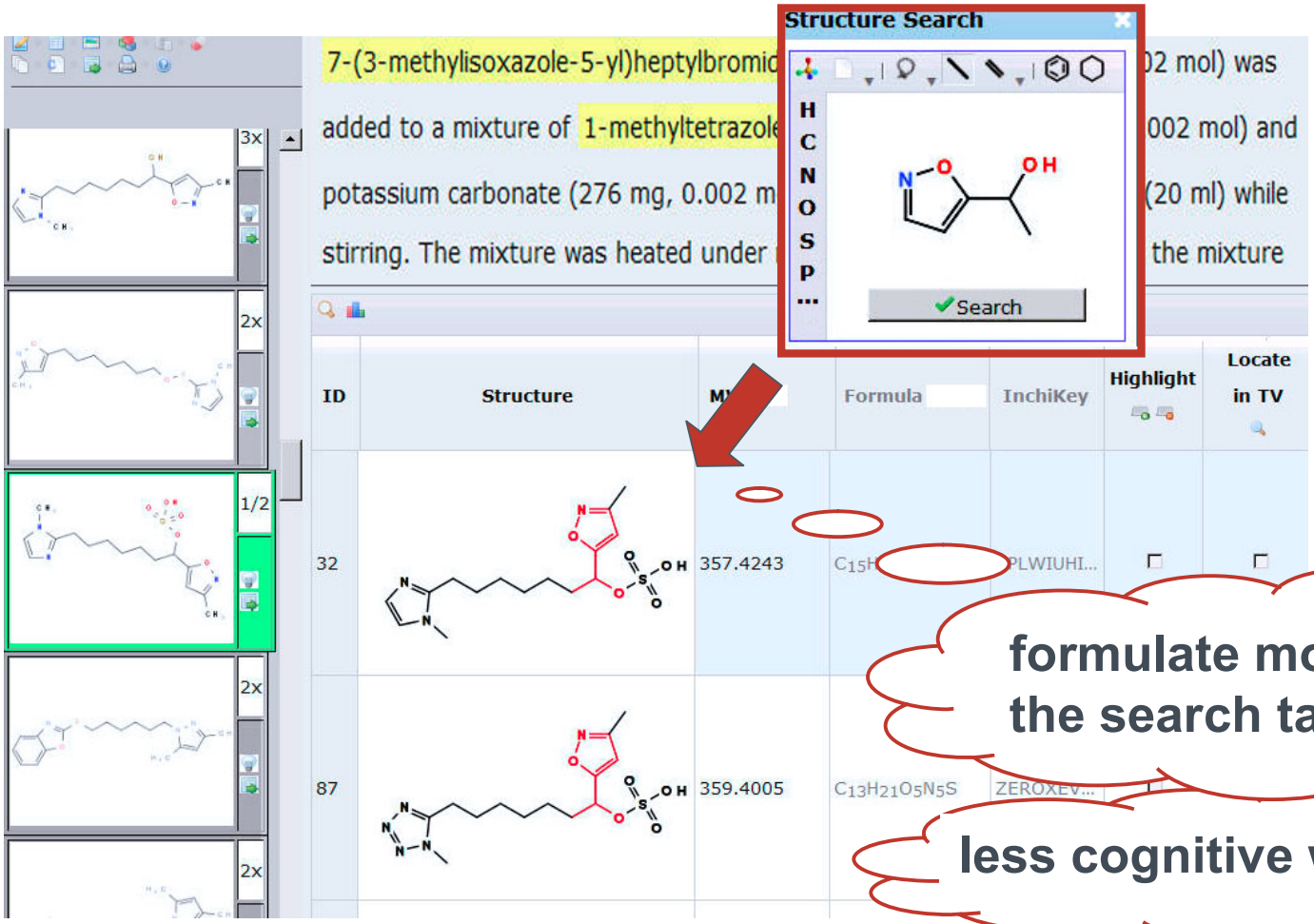




new technical features
(selectivity, potentcy, …)

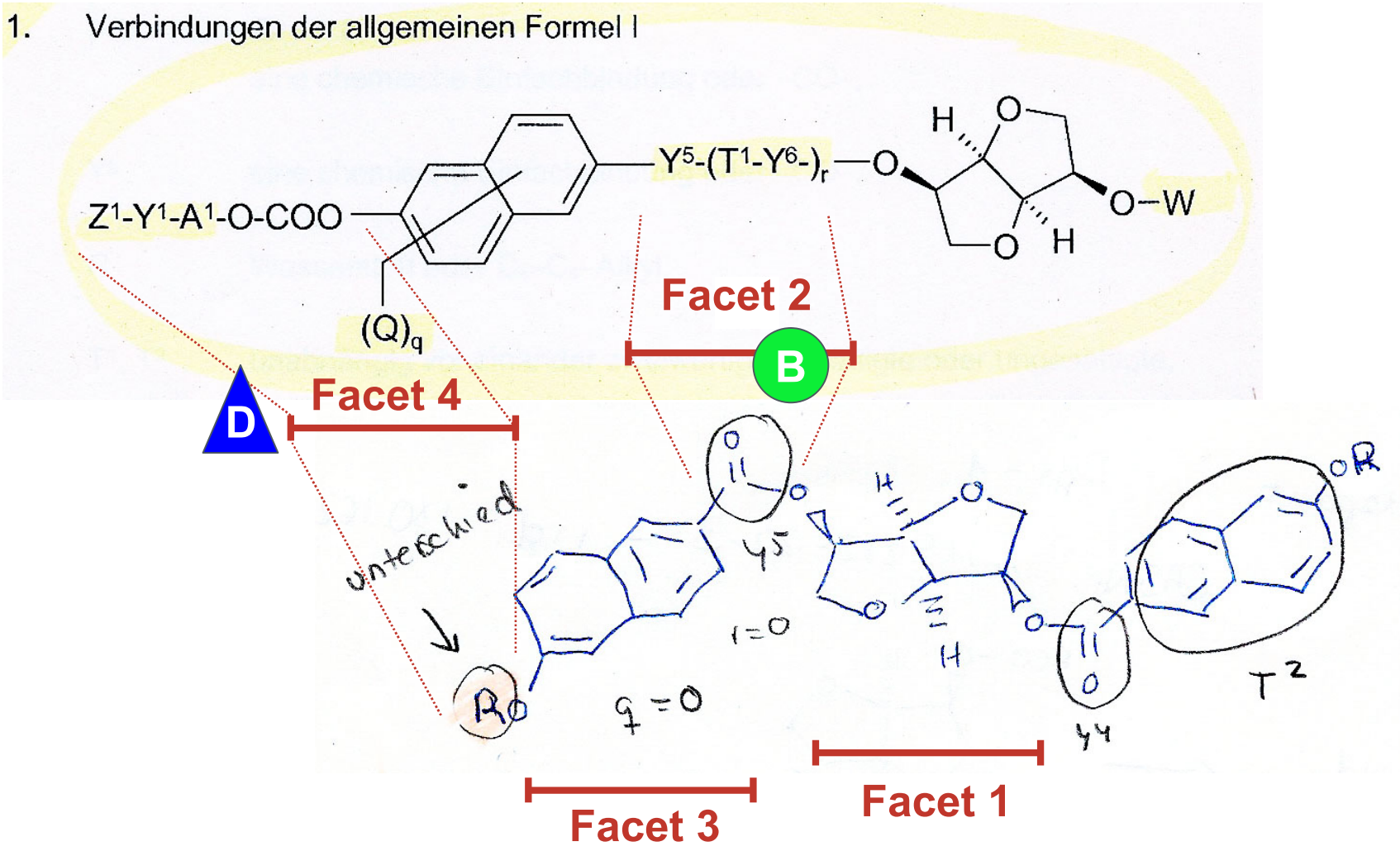
common technical features
(core structure, pharmacophore, …)

Tools to highlight distinguishing features



- Sub-structure search & visualisation tools in annotated document

Tools to search distinguishing features

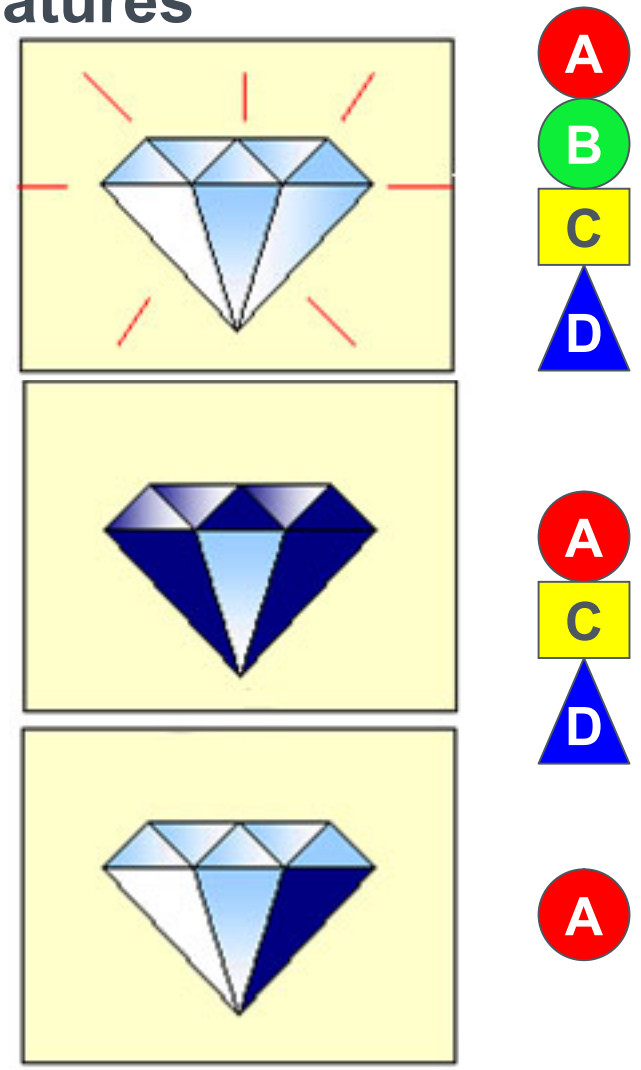


Tools to search distinguishing features

- Facet = technical concept or aspect of the invention

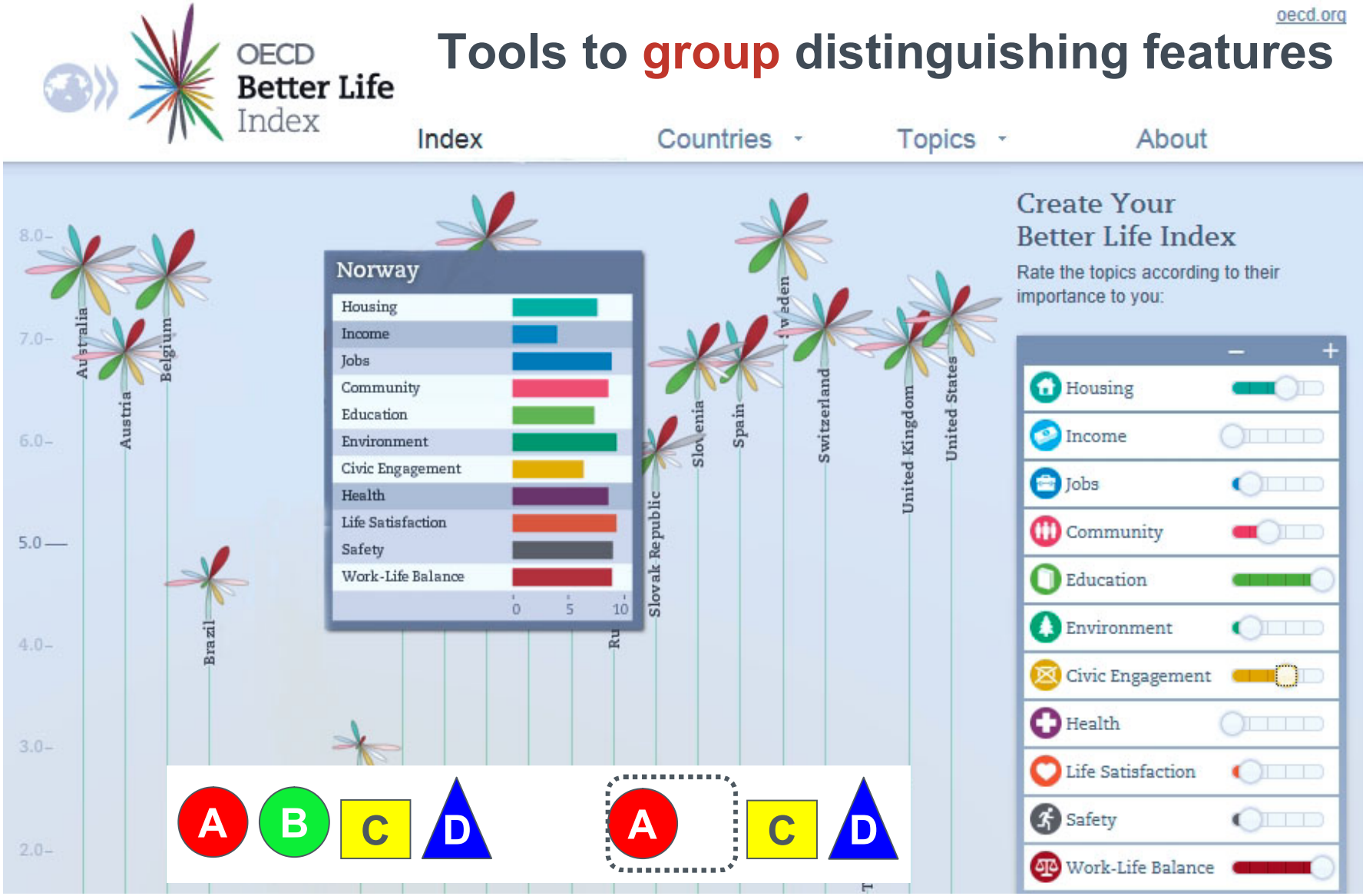
- Using “Boolean” search
  - A and B and C and D
  - 0 Docs → novel !
    → inventive ?

- Using “FACET” search
  - A;B;C;D
  - ..facet 4 (take four last SS)
  - ABCD,
    ABC, BCD, CDA, ABD,
    AB, CD, AC, AD,
    A, B, C, D



- **The more facets found in combination the closer the hit to the invention**
- **Variable preservation of connectivity of factets (“smart similarity”)**

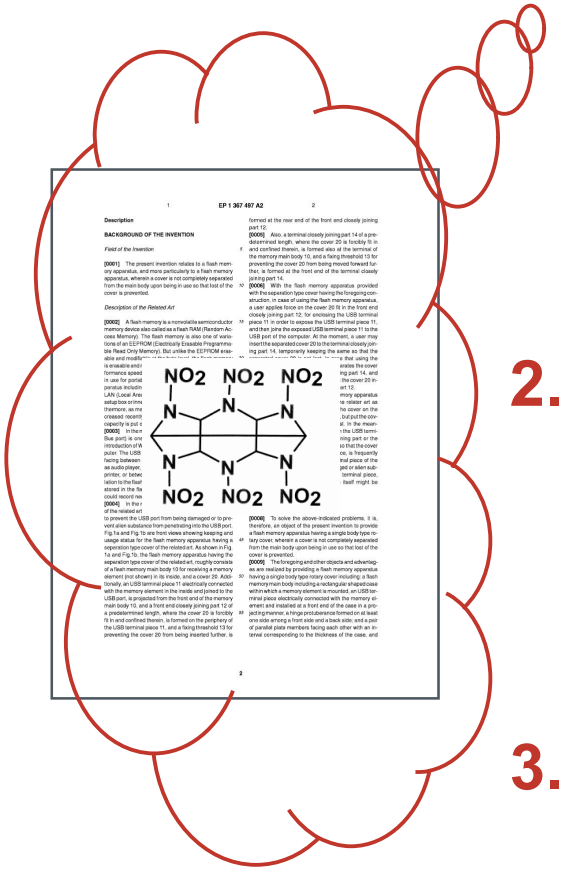
Tools to group distinguishing features



Source: www.oecd.org

- **Answer sets for detailed analysis (“smart aggregation”)**

Summary & Conclusions:



1. Mental Representations

- „heart & soul“ of human searches
- **explicit** (handwritten notes, drawings, record of search strategy, ...)
- **implictit** (logfile, user observations, …)

2. Chemical Annotation & Visualization

- a „must“
- assists spatio-visual ability
- not for comprehensive searching

3. Software

- support human way of searching
- smart „similarity & aggregation“ tools

(assist comprehension, concept development, search, …)

Thank you for your attention!

More information: Domenico Golzio (Dir. Search & Knowledge), Enrico Luzzatto (Dir. Pure and Applied Organic Chemistry)

Acknowledgement: Dr. Oliver Langer (ChemAnnotator), Sytse de Jonge (IT Roadmap Project Manager, Director)