HAL
open science

EnCOrE (Encyclopédie de Chimie Organique Electronique): an Original Way to Represent and Transfer Knowledge from Freshmen to Researchers in Organic Chemistry

Catherine Colaux-Castillo, Alain Krief

To cite this version:

Catherine Colaux-Castillo, Alain Krief. EnCOrE (Encyclopédie de Chimie Organique Electronique): an Original Way to Represent and Transfer Knowledge from Freshmen to Researchers in Organic Chemistry. 3rd International LeGE-WG Workshop: GRID Infrastructure to Support Future Technology Enhanced Learning, Berlin, Germany. 3 December, 2003, 2003, Berlin, Germany. pp.4. hal-00190462

HAL Id: hal-00190462

https://telearn.hal.science/hal-00190462v1

Submitted on 23 Nov 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

EnCOrE (Encyclopédie de Chimie Organique Electronique): an original way to represent and transfer Knowledge from freshmen to researchers in Organic Chemistry

Catherine Colaux-Castillo and Alain Krief
Facultés Universitaires N. D. de la Paix (FUNDP)
61, Rue de Bruxelles; B-5000 Namur; Belgique
Catherine.colaux@fundp.ac.be; alain.krief@fundp.ac.be

EnCOrE is an original proposal which is expected to allow to share and transfer knowledge in organic chemistry. The system will use MIDES a software, with a peer-to-peer architecture, which will allow to set up a technological and methodological frame to allow collaborative building of knowledge in between chemists. Learning GRID's services will help for Experimental electronic laboratory "LabCOrE" as well as for predictive computational tools.

Keywords: e-Learning, Grid, Social Informatics, Chemical Information technology, GRIDs for chemistry.

1. ORGANIC CHEMISTRY EXPERIMENTS AND REPRESENTATION OF DATA

Chemistry is the sciences of change: "La Chimie crée son objet" (Berthelot).[1] Matter is part of our day life and belongs to chemistry. Compounds as mixture or as pure products are found in nature or prepared de nuovo by chemical synthesis have often specific properties (dyes, pesticides, health care, material sciences, …).[2] Useful products which are present in nature but in poor quantities or are unknown, have to be prepared (synthesized) by chemists from less elaborated compounds. The choice of the starting material, the number of subsequent reactions which have to be used are determined by a strategic plan elaborated by the chemist according to his knowledge.[3] Use of less expensive ingredients and the shortest route, has a dramatic impact on the final cost of the product. In order to determine such plan, the chemist works by analogy and use rules which have been dictated by observations and physical principles. In fact the chemist works conceptually on molecules which are ideally *the smallest part of a compound having the properties of the whole compound.*[4] Those possess a structure, a design involving atoms and bonds, usually in three dimensions, which can be determined by several physical methods such as X-ray crystallography.[4] The structure consist of substructures which belong either to the skeleton (chains or cycles, hindrance constraints, …) or to the arrangement of atoms (functional group). Those confer specific physical and chemical behaviours to all the compounds which possess them.

Reactions imply one or several products which when mixed, in a close environment, spontaneously or under the action of applied energy (heat, light…) are transformed to one or several new compounds. The chemist represent these events as "chemical equations" in which the structure of the starting materials and products formed are respectively presented at the left or the right side of an arrow, which symbolizes the time elapsed, the human "actions" (heating, cooling, addition of a product, …) and the observations made during the chemical process.[4] Keywords related to the products, to the structures or to the reaction characteristics are linked to the "chemical equation".

Chemists think, work, represent, organise and transfer their knowledge in three main distinct ways: Experiments,[5,6] Chemical equations and texts linked to references.[7-10] Prediction using computational tools are now available and more and more used by experimentalists in organic chemistry.[11]

Databases involving structures (more than 17 millions) and "chemical equations" (several millions) are available from early 1900.[12] Electronic versions are available since 1985 and widely distributed since 2000.[13]

Basic knowledge is disclosed in manuals (textbooks)[4,7,8] and more elaborated or specific one are found in dedicated books,[9] series of books[6,10a] and encyclopaedias.[10b] In general those describe, in textual environment, mainly generic "chemical equations" and a few specific "chemical equations" with comments about their

generalisation, limitation, and rationalization. No media to display those data in an integrated format is up to now available.

2. ORGANIC CHEMISTS AND KNOWLEDGE IN ORGANIC CHEMISTRY

Organic chemists perceive their subject as intellectually highly structured, with many interconnected ideas. There are many quite different ways of teaching it, because there are many starting points, but the end result is often the same: a broad understanding and a shared but opaque language to outsiders. The sense of logic to the interconnected ideas disguises the fact that there is a serious problem in making the subject truly systematic. Classifying all the known organic reactions, for example, into conceptually recognizable groups sounds as though it would be easy, but in fact, apart from some straightforward seeming groups, like oxidation, reduction, addition and elimination, the categories are not well defined. Even some of those reactions pose problems.

Many concepts, very well accepted and used, cannot, as yet, be rendered into a format that computers can manipulate, although most organic chemists will be blind to the difficulties, and probably surprised how little can be done beyond the calculations, the handling of chemical structure databases and the word processing that they are familiar with. One of our aims is to try to reveal the underlying structure, and to achieve some systematic understanding of what organic chemists actually mean.

For this reason and others, this is an immensely ambitious project, for we are not simply entering the data to be found already in the secondary chemical literature, and providing an index to it. There are already more or less effective ways of searching large chemical databases like *Chemical Abstracts*[12a] and *Beilstein*.[12b] Instead, we shall collect afresh the fundamental chemical information needed, as one would if one were writing a book or series of books, and enter it into the encyclopaedia, in order to format it and structure it to suit more flexible computer-based retrieval. We expect that our finished product will allow students at all levels from first degree upwards to make connections that books and indexed databases are not designed for. We expect that it will contribute to the design of new ways of teaching, especially of organic chemistry, which has complex problems of having to handle words, chemical structures, mathematical formulae and experimental techniques, all equally important. We also expect that, by revealing connections that are not currently recognised, it will even expose areas of ignorance we are not aware of.

3. GOALS TO ACHIEVE AND ARCHITECTURE OF OUR PROGRAM

We want to achieve two goals related to Organic Chemistry and based to online transfer of knowledge:

3.1. Experimental electronic laboratory “LabCOrE”:

The objective is to teach experimental Organic Chemistry using exactly the same environment as the one of the most sophisticated research laboratories in universities or industries. This will be possible using a reaction flask which is connected to a robot,[14] like those used in combinatorial chemistry, and linked to the net. It is also necessary, for an optimal process, to link these robots to equipments which permit to follow the reactions, to isolate pure products (Chromatography: TLC, GC, HPLC) and to allow their structure determination.

Moreover this project will allow the "student" to adjust the conditions to get the best results. Background information, conceptual approach to the most adequate solution (ratio of starting materials, solvents, experimental conditions, temperature, and reaction time) will be helped by the concomitant use of EnCOrE.

3.2. Electronic Encyclopaedia EnCOrE

3.2.1 Architecture of EnCOrE

EnCOrE (Encyclopédie de Chimie Organique Electronique) tends to integrate and organize all the existing knowledge in organic chemistry. It should provide the average user all the information he needs. For that purpose we planned to manage interconnected huge databases of texts, chemical equations and references. The system will be built (Figure) according the conceptual view of the chemist (what do we want to include) and the need of the user (what and how the user wants to get the information).

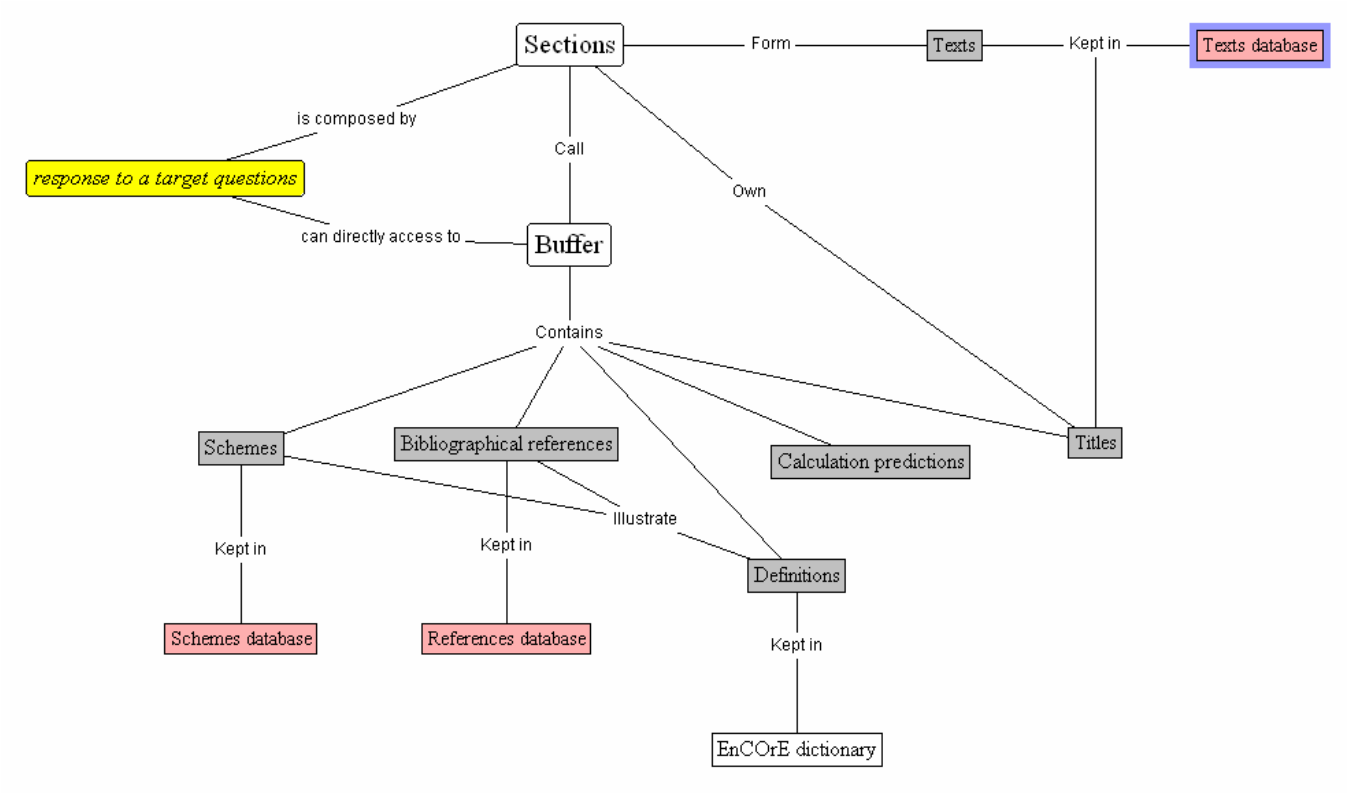

FIGURE 1: Architecture of EnCOre

The huge interconnected databases of texts, plans, chemical equations (Scheme database in Figure), references, and dictionaries, described above, will be linked to a “buffer zone” where all the information could be retrieved using integrated software's. These include predictive computational tools which require the Grid to work optimally (Figure). This kind of architecture will assure rapidity and efficiency of our system.

3.2.2 Unitex Software

We have selected Unitex software to manipulate, treat and search the texts.[15,16] It will help us to find most of the chemical words in our corpus using electronic standard dictionaries. These dictionaries are lists of words associated with coded grammatical indications which can be used by the syntax analyser. They will be employed to extract unrecognized chemical words which will be included in our electronic chemical dictionaries associated to grammatical and specific indications. Local grammars (graphs) should be therefore drawn particularly useful descriptor of "syntax motives". In a first approach, graphs will be used to cut each text into pertinent sections, later the same program should be employed as an efficient search engine.

3.2.3 IUPAC and MIDES Chemical Dictionaries

At the beginning we propose to use IUPAC (International Union of Pure and Applied Chemistry) dictionary which contains basic chemical definitions[17] to structure and analyze the texts, but we plan to use as soon as possible our own dictionary “MIDES electronic dictionary” which will describe the meaning of each word according to his each context in which it is used with adequate references and Schemes.[18,19]

For that purpose we will use MIDES (Modèle Interactif de Découverte Scientifique)[18,19] a software, which uses a peer-to-peer architecture, we are working on to “set up a technological and methodological frame to allow collaborative building of knowledge in between chemists”.

REFERENCES

[1] Berthelot, M. "La synthèse chimique", Librairie Germer Baillière et Cie, **1876**

[2] (a) Selinger, B. "Chemistry in the Marketplace" HarcourtBrace and Company, London, 1998 (b) Pybus, D. H.; Sell, C. S. "The Chemistry of Fragrances" Royal Society of Chemistry **1999** (c) Coultate, T. P. "Food" The Chemistry of its Components Royal Society of Chemistry, Fourth edition, **2002**.
[3] Corey, E.J.; Xue-Min Cheng "The logic of Chemical Synthesis" John Wiley and Sons, New York, **1989**.
[4] (a) Atkins, P. W. "Physical Chemistry", Oxford University Press, Oxford, Fift edition **1994** (b) Sykes P. "A guidebook to mechanism in organic chemistry" Longman, Fourth edition, **1975**.
[5] (a) Lowenthal, H. J. E. "A Guide for the Perplexed Organic Experimentalist" 2nd edition, John Wiley and Sons, New York, **1990** (b) Harwood, L. M.; Moody, C. M. "Experimental Chemistry, Principles and Practice Backwell Scientific Publications, Oxford, **1989**.
[6] Organic Syntheses Series John Wiley and Sons, New York.
[7] (a) March, J. "Advanced Organic Chemistry: Reactions, Mechanisms, and Structures" John Wiley and Sons, New York, Fourth edition, **1992** (b) Smith, M.A. " Organic Synthesis" McGraw-Hill, Inc., **1994**.
[8] Carey, F. A.; Sunberg, R. J. "Advanced Organic Chemistry" Part A: Structure and Mechanism; Part B: Reactions and Synthesis, Kluwer Academic/Plenum Publishers, New York, **2001**.
[9] (a) Davis, A. G. "Organotin Chemistry" VCH, Weinheim, **1997** (b) Tsuji, J. "Organic Synthesis with Palladium Compounds" Reactivity and Structure Concepts in Organic Chemistry 10, Springer-Verlag, Berlin, **1980** (c) Krief, A; Hevesi, L. "Organoselenium Chemistry I" Springer-Verlag, Berlin, **1988**.
[10] (a) Organic reactions Series Wiley and Sons, New York (b) Comprehensive Organic Chemistry, Pergamon Press, Ed. Trost, B.M.; Fleming, I. 1991.
[11] "Cache Software" from Fujitsu Company for example: http://www.cachesoftware.com/
[12] (a) Chemical abstracts, American Chemical Society, http://www.cas.org/ (b) Beilstein, Beilstein Institute, http://www.beilstein-institut.de
[13] (a) Scifinder and SciFinder Scholar http://www.cas.org/SCIFINDER/SCHOLAR/index.html (b) STN: http://www.cas.org/casdb.html (c) Crossfire for Beilstein MDL Information Systems GmbH, http://www.beilstein.com/beilst_2.shtml.
[14] Robots for parallel synthesis. (a) Chempseed: http://www.chemspeed.com/VLT100.html (b) ProSense Laboratory & Process equipment: http://www.prosense.net/
[15] Gross, M. *Ann. Télécommun.*, 5, *44*, n°1-2, **1989**.
[16] http://www-igm.univ-mlv.fr/~unitex/
[17] IUPAC Compendium of Chemical Terminology: http://www.chemsoc.org/chembytes/goldbook/
[18] Paraguaçu, F.; Cerri, S. A.; Costa, C; Untersteller, E.; IEEE/ITRE2003, August 10-13, **2003**, Newark, New Jersey, USA.
[19] F. Paraguaçu, E. Untersteller (2003) Interactive Model for Scientific Discovery Application to the Conception and Realisation of a Dictionary in Organic Chemistry. Technical Report LIRMM Univ. Montpellier II, ICSN CNRS, FUNDP, and FNRS. May 15th, 2003.

ACKNOWLEDGEMENTS

This work has been supported by the FNRS (Fonds National de la recherche Scientifique, Bruxelles, Belgique), by the CNRS (Centre de la Recherche Scientifique, Paris, France). The authors thank Prof. I. Fleming (Cambridge University, United Kingdom), Dr E. Untersteller (Institut de Chimie des Substances Naturelles ICSN-CNRS, Gif sur Yvette France and FUNDP, Belgium), Ms M.A.Marchal (Facultés Universitaires Notre Dame de la Paix, Namur, Belgium, and Université Catholique de Louvain, Louvain la neuve, Belgique) for their helpful contribution.
Encore is a collaborative project involving five active sites: (a) Facultés Universitaires Notre Dame de la Paix (FUNDP), Namur, Belgium, (b) Institut de Chimie des Substances Naturelles ICSN-CNRS, Gif sur Yvette, France) (c) Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (Lirrm), Montpellier, France (d) Ecole nationale Superieure de Chimie de Montpellier (ENSCM), Montpellier, France (e) Université Catholique de Louvain, Louvain la neuve, Belgique).