# Journal of Applied Sciences

ISSN 1812-5654

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 5 (6): 1062-1065, 2005
ISSN 1812-5654

# Flexible Structural DBMS Case of Chemical Study

A. Chouarfia, B. Messabih and H. Bouziane
Institut d'Informatique de l'Université USTO, BP 1505, Oran M'Naouer Algérie

**Abstract:** DBMS systems and databases in chemistry, whether for retrieval information or retrieval strategy, are usually limited by their procedural reasoning tools. The limitations of usual systems are often caused by a lack of flexibility due to rigid structuration of the search space, the choice of given logic and very often, the fuzzy nature of some of the basic state of space blocks. We are currently investigating the power of analogical search and retrieval tools with an eye to developing a flexible system through the case memory strategy and the use of case memory storing potential similar features chosen by users. The context adaptability parameters, topology, metrics and physical properties.

**Key words:** Artificial intelligence, machine learning, analogy reasoning case based reasoning substructural search, system, fuzzy search, similarity metrics, DARC system

## INTRODUCTION

The concept of an adaptive and flexible system, capable of internal reorganization according to the problems to be solved, was oriented by our knowledge of an original system of molecular chemistry developed from the DARC system and producing data banks and various expert systems. We note that structural databases today constitute the heart of chemical information systems. These were developed by successive generations. The first generation was described by fragmented models and handled by low level languages (logical combination of fragments). It was an application of engineering to documentation techniques. Because they divide compounds into fragments to adapt them to computing handling, these models distort and reduce the semantics of the structural information. The second generation, based on the DARC topological model proposed by Dubois had a more intensive scientific development. Chemical structures are here perceived as oriented graphs implemented as tables and handled by graphic language. This generation, with its important scientific activity, was industrialized (EURECAS and CAS). Today, these systems currently include 11 million compounds available on line by substructure. Moreover, certain systems of patent banks also derive from it. These highly performing systems endowed with generic tools, primarily use procedural reasoning. Most of the derived molecular expert systems, are only partially linked to these large databases oriented and optimized for well defined documentary requirements. To go beyond their limitations, it is well known that new systems must dispose of a multiform reasoning mechanism.

**Multiform reasoning and data handling:** Like the rich variety of human thought mechanisms, there are several types of reasoning in Artificial Intelligence. The main problem involved in formalizing reasoning is to succeed in conducting a series of operations on symbolic entities while preserving their veracity and the coherence of deduced conclusions. Chemistry is privileged by its symbolism of structures and families. Mathematical logic is also a perfect example wherein formulation is an essential aspect. Roughly, logical reasoning can be divided into three categories:deductive reasoning in which valid conclusions are deduced from a set of promises, inductive reasoning where, on the contrary, premises are generalized and addictive reasoning where plausible reasons are linked to a set of premises often based on limited relations. Many languages and systems are founded on deductive reasoning. This type of logic has serious advantages for building a knowledge based system[1,2]. The reasoning rules are syntactically and semantically defined; they are moreover simple and easy to use. However, this logic contains intrinsic limitations that prevent it from efficiently dealing with problems of great importance such as analogy, the uncertain, dealing with time, questioning conclusions, etc.

---

**Corresponding Author:** Dr. A. Chouarfia, Institut d'Informatique de l'Université USTO, BP 1505, Oran M'Naouer Algérie
E-mail: chouarfia@mail.univ-usto.dz

**Learning by analogy:** Granted the technical premature nature of integrating all these modes of reasoning in one system, we propose in our study to integrate reasoning by analogy whose application aspects will be those of a search system of chemical information where analogy searches are carried out on the level of substructures. Present reasoning by analogy makes use of prior cases and can thereby be likened to learning techniques, even if our ambition is for analysis to use and integrate different logics (deduction, induction, abduction).

**Reasoning by analogy and information search systems:** Analogy plays an important part in human daily and professional activities: it intervenes in learning, in decision making, argumentation, explanation, understanding language, etc. More specifically, it is closely related to the processes of memorization and the use of stored information. For this reason, analogy can be a source of inspiration and of carrying out intelligent systems for managing large data and knowledge bases. It can be used in two ways: to help a user to find the knowledge corresponding to his need and/or to help improve the quality of the data base management system. Case-based Reasoning, derived from analogy reasoning, is a promising approach for the practical realization of such systems. The idea is to start from previously encountered experiences and cases in order to solve new problems. In case-based reasoning, the handling of a new problem is intuitively linked to the learning mechanism, thereby making it possible to assimilate this problem and to work out a learning mechanism for some later use[3,4]. Figure 1 shows the process of case-based reasoning and learning.

**Assign index:** The characteristics of a new problem are used as descriptive indices of this problem.

**Retrieve:** The indices and similarity metrics are used to retrieve in the case memory for one similar to the present case. In chemistry, indices can belong to the structural display that symbolizes and expresses numerous analogies and associated metric data.

**Modification:** The past case is modified to include conditions belonging to the current situation, thereby leading to a proposed solution.

**Test explain, repair:** The proposed solution is tested to check its adequacy to the problem.

If the test is satisfying, the solution is given to the user, then filed in the case memory. If the test is negative, the cause or causes of the failure must be found and eliminated before retesting.

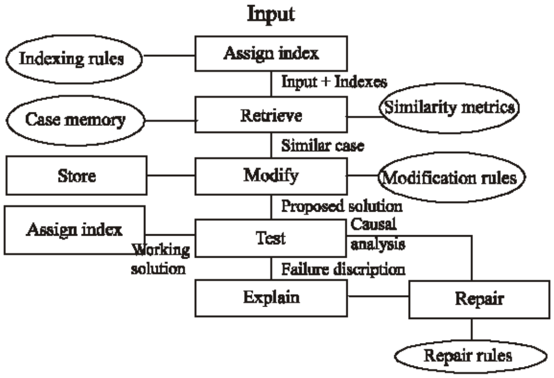

Fig. 1: Process of case-based reasoning

Studying this process highlights the position of analogy on the problem level between case-based reasoning and structural database management systems[5]. These problems involve indexing information, formulating queries so as to obtain relevant information, similarity metrics, etc. In addition to this beneficial analogy, case-based reasoning enables a flexible and evolving structural DBMS to gradually acquire experience and to become adaptable to the needs of a specific user.

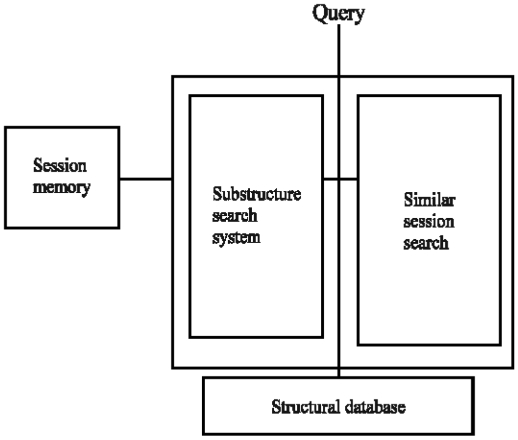

Fig. 2: System architecture

According to this architecture (Fig. 2), the substructural search system uses the case-based reasoning much as a chemist would use his own experience to answer a query more satisfactorily[6]. By adding a component part of case-based reasoning (capable of learning) to a substructure search system, we want to achieve a double objective:

i) To adapt a search to a use context by exploiting both the successes and the failures of previous sessions stored in the case memory.
ii) To achieve a long term enrichment of the system know-how by using new knowledge that then be extracted from the session memory.

**System architecture and functioning:** The solution we have in mind consists in first elucidating the various possible choices during each phase of the search process (Fig. 3), then adjusting them gradually and progressively all during the use of the system[7].

We note that a chemical substructure search system uses conceptual and symbolic expressions of structural knowledge. It is, therefore, a very favorable area of application for testing the ideas evoked below.

**Interpretation:** For a substructure search system, to interpret means to select from the database the set of structures with a given structural constraint. This binary restriction: a structural constraint, verified or not, often obliges the chemist to adapt this his queries, thereby rendering the substructure search system rigidly unintelligent. Perhaps in a search implying a given constraint the user would be ready to accept some slight differences or even a slightly greater difference without being able to specify the limits of this advance what specific criteria of acceptance he can use. He would like the system to suggest those candidates likely to be accepted and then to let him choose among them from such experienced man aging of a substructure search system.

**Sorting:** The selected structures are sorted according to their similarity with the query. The literature proposes many methods for calculating similarity coefficients between a query Q and a structure S. The results of these calculations can be incorporated into the database or in annex as learning elements. According to the case, the similarities are based either on the degree of inclusion of a query Q in the structure S (by identification of the maximal subgraph) or else on the number of structural fragments that the query and the structure have in common. These measures can be combined to allow for several classifications.

**Display:** It is during this phase that the choice is made of the number of structures to propose for display and to the users judgment. There are two possible approaches: either there is a threshold and only those structures whose similarity with the query is greater than this threshold are displayed, or the user is asked to set a maximum of structures, that he accepts, to be visualized

**Session memory and adapter context:** This is the evaluation stage, with the users help, of the relevance of a search session. Such relevance can be tested by the precision coefficient defined as the number of relevant structures proposed as compared with the total number of relevant structures proposed, or by some heuristic methods, i.e. the users level of satisfaction with regard to the proposed structures. This stage is very important, for the system learns to improve its performance from one session to another. The great similarity between the characteristics of a search session and a structure led us to propose a session memory organization (case based) identical with structure base structuration. Just like a structure, a session is characterized by an index a content. The indices include, in addition to the primitives, a session relevance index evaluated at the end of each session. The content comprises elements relative to the different search stages, the choices made, the strategy adopted and the list of the chosen compounds.

The search for a session similar to the present is first carried out in the subset of non relevant sessions (those having led to failure), in order to anticipate and prevent this failure. This order of priority, that we can all learning from failures corresponds to the first objective we want for our system: to improve search strategy and adapt it to a use context[8].

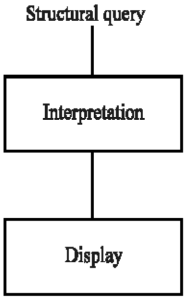

Fig. 3a: Classical search procedure

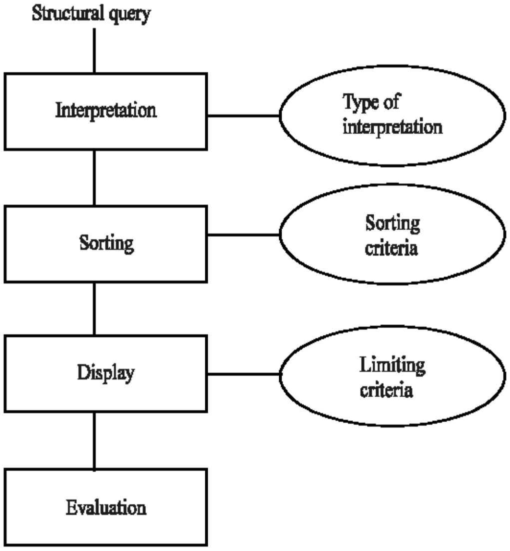

Fig. 3b: Parametric search procedure

## CONCLUSIONS

In this study we describe the main of outlines of an approach combining case based reasoning and fuzzy substructure search, the aim of which in to improve the search for information thanks to the use of a session memory. When this memory will be judged sufficiently abundant, we shall have to define a judicious utilization of this experience bas so as to enrich the system both in knowledge and in performance. The proximity relations between cases and the metaknowledge that we can excerpt from this memory will enable us to enrich the systems potentialities by browsing tools furnishing reasoned help to the user. With these tools, our system is characterized by openness and flexibility and can generate profile files adapted to specific user needs and conduct chemical diversity searches useful for data mining actions and chemical knowledge creation. These form part of strategies for holding library information. The prototype currently being developed[9], is built around a fuzzy substructure search system based on the latest DARC topological approach which is currently operational in a UNIX environment on a IBM RISC 6000 graphical station.

## REFERENCES

1. Forsyth, R. and Rada, 1986. Machine Learning: Application in Expert Systems and in Information Retrieval. Ellis Harwood Series in AI.
2. Chiamella, Y.B., 1987. A prototype of an intelligent system for information retrieval. IOTA, Information Processing and Management System, 23: 285-303.
3. Slade, S., 1991. Casebased reasoning: A Research paradigm. Al Magazine 12: 42-53.
4. Chen, Z., 1994. Enhancing database management to knowledge base management system. The Role of Information Retrieval Technology, Information Processing and Management System, 30: 419-453.
5. Rouvry, D.H., 1994. The necessity of analogics in the development of science. J. Chem. Inform. Sci., 34: 446-452.
6. Golging, A.R. and P.S. Rosenbloum, 1991. Improving roule-based systems through case-based reasoning. In: Proc. Am. Assoc. Artifical Intelligence Conf., Anaheim, CA, pp: 22-27.
7. Messabih, B., 1997. Fonctionalité dun SGBD Structural 2D/3D orienté vers la conception assistée. Ph.D Thesis, University Paris VII.
8. Chouarfia, A., B. Messabih and H. Chouarfia, 2000. Intelligent structured databases management: Case of Chemical structures retrieval. Scientific Research Outlook in the Arab World and the New Millennium Science and Technology. University of Sharjah, UAE.
9. Work Done by SYCOGMA, SA Paris.