Organic Qualitative Expert System

Mario Garcia, Rajan B Gujjarlapudi, Kaushik Hatti, and Carl Steidley

Department of Computing and Mathematical Sciences
Texas A&M University Corpus Christi
garciam@falcon.tamucc.edu

Abstract

The idea of using a computer program to simulate a human expert (*i.e.* an informant) in the Qualitative analysis of Organic compounds has been introduced lately and it is gaining popularity with the students day by day. This expert system does not intend to replace hands-on laboratory work but it does allow the student to work on a greater number of samples in a short time and allows development of problem-solving skills and development of strategies for dealing with qualitative analysis. This system should be used simultaneously with the procedures and the results obtained from lab and hence reducing the work for the student.

Introduction

One of the greatest challenges to the chemist is identifying the substances that are obtained from chemical reactions or isolated from natural sources. Although structure elucidation may be difficult and time consuming, the availability of modern spectroscopic techniques in combination with chemical methods has greatly facilitated the aspect of Organic Qualitative Analysis. The classical approach to structure elucidation involves converting an unknown liquid or solid into a solid which is known as Qualitative derivative Analysis.

Qualitative analysis is the analysis and identification of unknown organic compounds, which constitutes a very important aspect of experimental organic chemistry. There is no definite set procedure that can be generally applied to organic qualitative analysis. This system has been designed for use by the students taking the Organic Lab at Texas A& M University-Corpus Christi. The students are given 4 chemical compounds out of which the chemical formula is given for one compound, molecular weight for one, functional group for another. No data is given for the fourth unknown compound. The unknown will contain only one type of functional group. The type of functional group of the unknown will be limited to one of the following: Amines, Aldehydes, Ketones, Carboxylic acids, Halides, Alcohols and Aromatic compounds. The report offers a discussion of a prototype and the way they are constructed, along with an expert system that is used to help solve the Qualitative Analysis of Organic Compounds. This program tries to determine the possible tests on the unknowns and gives a list of compounds.

Copyright © 2004, American Society for Engineering Education

Literature review

Most of the literature required for the project, books and journals were suggested by the expert, in this case Mr. Kaushik Hatti working as a Teaching Assistant for the Organic lab at Texas A& M University- Corpus Christi. The most useful book in this project was The *Experimental Organic Chemistry* by *Gilbert and Martin* [5]. Other books which were helpful in this case were Experimental Organic Chemistry by Daniel R. Palleros. Many important points were noted down from the *Organic Chemistry Laboratory Manual* by *Dr. Mark C. Morvant.*[10]

Most of critical information required for this project was obtained from *Gilbert and Martin* [5] which addresses a critical need faced by researchers in all fields of the human sciences - how to draw valid meaning from qualitative data. It provided methods of analysis that are practical, credible and reliable. The authors refer to many ideas and references that draw on the experience of the authors and many colleagues in the design, testing and use of qualitative data analysis methods. Each method of data display and analysis is described and illustrated in detail, with practical suggestions for adaptation and use. The growth of computer use in qualitative analysis is reflected throughout this volume, which also includes an extensive appendix on criteria useful for choosing among the currently available analysis packages.

There have not been many expert systems developed for qualitative analysis. One of them is the ORQA (Organic Qualitative Analysis) which helps students identify organic unknowns in the laboratory. It acts as a private tutor for the students. This system has been developed using the Side-Eight software. It contains a list of compounds which are not been prominently used in the university. Our expert system has been developed using CLIPS. The list of compounds used have been specifically designed to meet the requirements of the students taking the Organic chemistry lab course at the university, but the compound list could be incremented to accommodate the requirements of other universities if needed.

History

Chemistry has been called the science of what things are. Its intent is the exploration of the nature of the materials that fabricate our physical environment, why they hold the different properties that depict them, how their atomic structure may be fathomed, and how they may be manipulated and changed. Although organic reactions have been conducted by man since the discovery of fire, the science of Organic chemistry did not develop until the turn of the eighteenth century, mainly in France at first, then in Germany, later on in England. By far the largest variety of materials that bombard us are made up of organic elements. The beginning of the Ninetieth century was also the dawn of chemistry, all organic substances were understood as all being materials produced by living organisms: wood, bone, cloth, food, medicines, and the complex substances that configure the human body. Because of the human's wonder of natural life, organic materials were believed to possess an enigmatic "Vital Force." Thus organic chemistry was separated from inorganic chemistry, and it became its own field of science. By the turn of the Nineteenth the "Vital Force" theory was immensely discredited, but this branch of science still stayed separated from inorganic chemistry.

Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition
Copyright © 2004, American Society for Engineering Education

In 1845 Kolbe synthesized acetic acid, the chief component in vinegar, in a flow of reactions starting with Carbon, the experiment is demonstrated better defined since acetic acid ($C6H4O2$) is a carbon-carbon bond. The theory of vitalism, like many other scientific theories, disappeared slowly under the weight of accumulated evidence rather than as a consequence of any one brilliant and enlightening experiment. Structural theory, which developed in the 1860's, started the second major period of growth in the organic chemistry field. The development of a detailed picture, by using pure reasoning of both atomic organization and the shapes of molecules stands as a great milestone of the development of human intellect. At almost the same point in time, Kekule in Germany, and Couper of Scotland suggested that atoms in molecules are fused together by bonds. Their theory was that every atom is distinguished by having the same number of bond availability or valence number, where ever that particular atom appears in any compound. The main notability of organic compounds is having strong carbon to carbon bonds.

This was recognized in the theory, and was used to help understand large molecules, possessing many bonded carbon atoms. Carbon is the cement that holds their molecules together. So far, this theory has gone through rigorous testing, and has not been proven inadequate to this day, as of now it is a law. Kekule and Couper's theory was not all without fault; it is surprising that they did not recognize atoms as three-dimensional objects if they were to be understood as true particles of matter in space. It was not until 1875 when van't Hoff and LeBel proposed their hypothesis of compounds and atoms taking up space. Their hypothesis went as follows: Four bonds of carbon were located at equal angles to each other in space, this would be a rectangular tetrahedron. The structural theory is not only a focal point of organic chemistry, but an amazingly simple idea. It states that by grasping that each carbon atom to form four bonds, tetrahedrally arranged in space, we are able to map the architecture of even the most complex molecules. The power of the theory is demonstrated by the statement that there has been no chemical observation that cannot be basically understood by structural theory.

Finally, although structural logic is extremely rigorous, it involves no mathematics. Unlike most sciences of equal complexity, much of organic chemistry is conducted without the use of formal math beyond elementary levels. The third and presently used theory in the history of organic chemistry ends with the description of chemical bonds as electron pairs, Lewis came up with this in 1917. Although a great amount of chemical reactions were already known and in active use to synthesize organic compounds into other compounds, only with this understanding of the nature of a chemical bond did a clear reason of the nature and mechanism of chemical reactions begin to appear. Thus if the nineteenth century was devoted to unraveling the fixed structures of molecules, the twentieth century will be devoted to the study of their transformations and isolating new compounds which can be done only with *Qualitative Analysis*.

In recent years the development of Chromatographic methods of separation and structural analysis by spectroscopic techniques have revolutionized the laboratory practice of organic chemistry. Nevertheless, interest in classic qualitative analysis remains high because it is an effective and interesting means of understanding fundamental organic chemistry. The classical system of qualitative analysis consists of six steps as shown in figure 1. [4]

1. Preliminary examination of physical and chemical characteristics.
2. Determining physical constants.

Copyright © 2004, American Society for Engineering Education

3. Elemental analysis to determine the presence of elements other than carbon, hydrogen and oxygen.
4. Solubility tests in water, dilute acid and dilute base.
5. Functional group analysis using classification tests.
6. Derivatization.

It is a tribute to the power of the system that one can identify an unknown organic compound with certainty, even though it may be one of several million known compounds. This expert system uses the novel feature of qualitative analysis which is that the negative results may be as useful as positive ones in the quest to reveal the identity of an unknown.

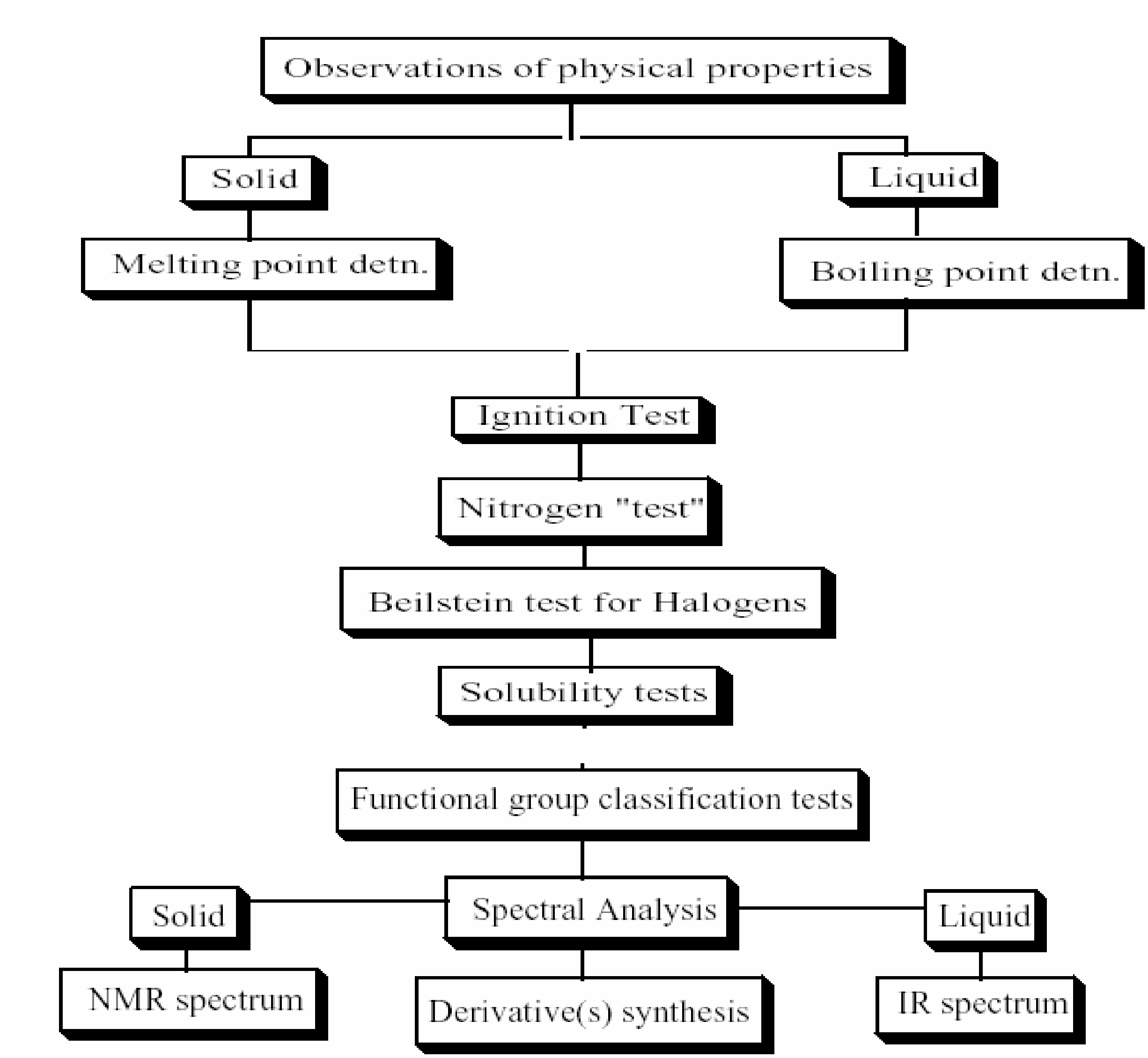

Figure 1 Qualitative Analysis steps

Copyright © 2004, American Society for Engineering Education

Expert System Architecture

An expert system, by definition, [4] refers to programs whose competence at a task derives from knowledge about the task domain, or knowledge base. A knowledge base is a collection of knowledge expressed using some formal knowledge representation language. Furthermore, knowledge is the objects, concepts, and relationships that are assumed to exist in some area of interest. Figure 2 illustrates the basic concept of an expert system ES. The KBS consists of two main components. The knowledge base contains the knowledge with which the inference engine draws conclusions. [8] These conclusions are the results of the user queries.

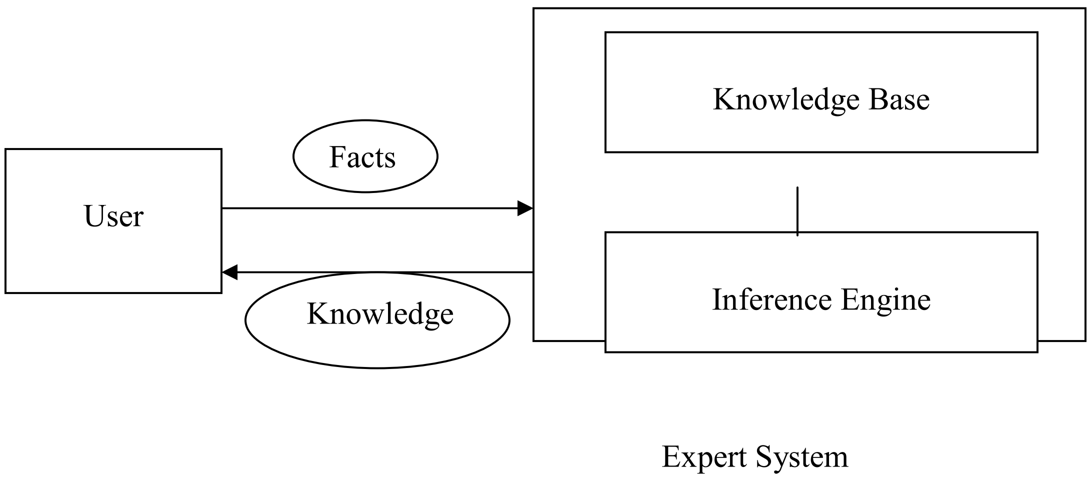

Figure 2 Basic Concept of an Expert System

Knowledge Acquisition

The objective of the knowledge definition stage is to define the knowledge requirements of the system. The knowledge definition stage consists of two main tasks: knowledge source identification and selection and knowledge acquisition, analysis, and extraction.[9] The main objective of these tasks is to produce and verify the knowledge required by the system in preparation for the next stage of knowledge design.

In order to be useful in the real world, expert systems must be able to solve complex problems in well-defined domains. To this aim, they must possess a sufficient amount of knowledge; hence the necessity to place particular interest on the problem of the acquisition of such knowledge, that is still the bottle-neck that prevents the large-scale production and diffusion of expert systems. Knowledge acquisition is the most important aspect of the expert system development. It is referred to as the process of getting and transforming appropriate information out of an experts head, document or any source into some manageable form.

Interviewing Experts: This involves the knowledge engineer having a face to face interview with the expert. The project team had discussions with number of former students who have taken the course to gather information about how the students wanted the expert system to be designed.

Copyright © 2004, American Society for Engineering Education

Books and Journals and Internet sources: Since Qualitative analysis is a broad concept there were many books used to do the knowledge acquisition. [1,2,3, 7,11, 13,14] Some of the new developments were made available in the form of journals. Most of the physical properties, flow chart and some of the derivative tests were obtained from the Internet sources.

Knowledge Organization

The objective of the knowledge design stage is to produce the detailed design for the system. There are two main tasks within this stage: knowledge definition and detailed design. [4] The knowledge definition task will involve representing the knowledge in the form of facts and rules using Clips [13] syntax. These rules were used to represent the different steps in a procedure and use the facts to make decisions.

This huge amount of information was organized in a simple way using the Inxight tree [6] software tool. The root node in this system was named as Main Menu. It has four children depending on the data available to the student. Three of the compounds have chemical formula, Molecular weight and Functional group respectively. The fourth unknown does not have any kind of data. The flammability test, functional group tests and physical properties succeed as children to the unknown compound.

Knowledge Representation

Knowledge representation is concerned with designing and storing knowledge that are known as facts and rules about a particular topic. [9] The elements of a ES typically consist of the following [4]: user interface – the mechanism by which the user and the system communicate; explanation facility – explains the reasoning of the system to a user; working memory – a knowledge base; inference engine – makes inferences by deciding which step to execute; agenda – a prioritized list of actions created by the inference engine; and a knowledge acquisition facility – a way to allow the user to enter information into the system and is most often an optional feature in a ES.

Implementation

The Expert System prototype was developed after organizing the knowledge acquired. Star Tree Studio™ 3.0 is a Windows application that uses wizards and an intuitive interface to walk you through the process of creating, designing and publishing Star Trees. Star Tree Studio was used to represent the knowledge base. Inxight tree [6] helped in the development of the rules for the system. It had a step by step description of the whole process of Qualitative Analysis. The three compounds which had information about them were first developed up to the level of physical property of a compound. Then all the children of the unknown compound were added one by one depending on the functional group to guide the user about all tests possible. This systematic approach to knowledge representation created an environment with no confusion and chaos.

The students were given 4 chemical compounds out of which the chemical formula is given for one compound, molecular weight for one, functional group for another. No data is given for the fourth unknown compound. The unknown will contain only one type of functional group. The type of functional group of the unknown will be limited to one of the following: Amines,

Copyright © 2004, American Society for Engineering Education

Aldehydes, Ketones, Carboxylic acids, Halides, Alcohols and Aromatic compounds. The report offers a discussion of a prototype and the way they are constructed, along with an expert system that is used to help solve the Qualitative Analysis of Organic Compounds. This program tries to determine the possible tests on the unknowns and gives a list of compounds.

All the rules were implemented using one of the popular tools for expert systems i.e., CLIPS. It is a *tool* that is designed to make the development of software to model human expertise easier. It represents (human) knowledge in three ways:

1) Rules for experience-based, heuristic knowledge
2) Deffunctions and generic functions for procedural knowledge
3) OO programming, also for procedural knowledge

The main reason for choosing CLIPS lies in its principal advantage as an efficient and flexible rule interpreter, but has functional and object-oriented constructs also, making it an excellent development language. Compared with compiled C++, the development cycle is much faster, with application changes being apparent within seconds of changing source code.

Since Qualitative Analysis requires the students to be in a constant conversation with the expert, it is most ideal for implementation of an expert system by replacing the human with an expert system. Since CLIPS is a rule base system, rules are used to represent heuristics, or "rules of thumb," which specify a set of actions to be performed for a given situation at any point during Qualitative analysis. Following are some examples of the rules used in the expert system

```
 (defrule rule3
        (gen c4h10o|C4H10O)
        =>
        (printout t "The following compounds match the formula you entered " crlf)
        (printout t "              Butanol-1          Press 1 to see the physical properities" crlf)
        (printout t "              Iso-Butanol        Press 2 to see the physical properities" crlf)
        (printout t "              Di ethyl Ether     Press 3 to see the physical properities" crlf)
        (assert (take(read)))

(defrule rule25
        (mol 137)
        =>
        (printout t "These are the compounds which match the Molecular Weight you entered ")
        (printout t "     2-Bromo 1-methylpropane      Press 1 to see the physical properties" crlf)
        (printout t "     2-Bromo 2-methylpropane      Press 2 to see the physical properties" crlf)
        (printout t "              2-Bromobutane       Press 3 to see the physical properties" crlf)
        (assert (temp1(read)))
)

(defrule rule124
        (temp84 3)
        =>
        (printout t "   Di Bromo Iodide " crlf)
        (printout t "   Boiling Point  58 c" crlf)
```


Copyright © 2004, American Society for Engineering Education

(printout t "   Density   0.009" crlf)
(printout t "   Color   Brownish" crlf)
(printout t "   Smell   Urea smell" crlf))

Testing

Once the expert system was developed, it was submitted to vigorous testing by the expert to detect any flaws. It was also tested by some of the former students who commented that the system reduced the work-load of the student drastically, which was the primary objective of developing the system. It was also tested by some of the other teaching assistants of the organic chemistry lab at the Texas A&M University-Corpus Christi who found the system to be an excellent guide for the students. They suggested including the safety aspects of the compound to be listed along with the other properties. An example of the introductory screen presented to the user is shown in figure 3.

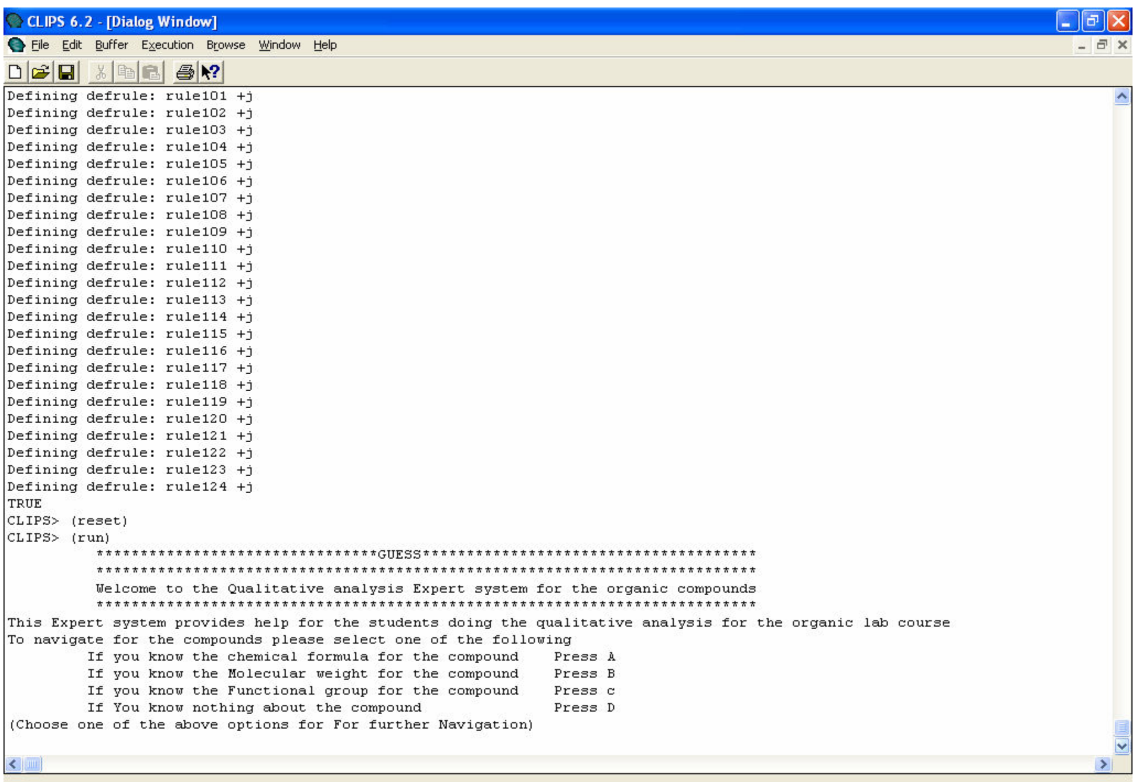

Figure 3. Example of the introductory screen in CLIPS

Conclusions

The potential benefits of expert system are to bring expert knowledge within the reach of less experienced engineers, thus reducing the need for an experience professional in this area to attend to every step of analysis. Also, this system would save considerable amount of man hours. There are many benefits that can be attributed to the use of an ES. These include improvements

Copyright © 2004, American Society for Engineering Education

in productivity, preservation of knowledge, economical benefits, improved quality of goods and services, training, and job enrichment. Of these training is the focus of the proposed system which inherently improves all other aspects. Improvements of training employees using an ES are that it establishes a "consistency in decision-making…better compliance to organizational policies and procedures…reduces training time…quickly improves skills…and assist in the documentation of decision-making information for future use and training." [8] An ES, when implemented, has the obvious potential to save money which traditionally is what drives product development.

With this research it was demonstrated that the use of expert systems can be applied successfully in different areas of science like Chemistry. The students who used the expert system were very satisfied with results. The main reason for embarking on developing this system was to make the job of the student easier. With little modification this system could be used in most of the universities where the pattern of unknown analysis is same.

Bibliography

[1] Burke S, and Danheiser R. Oxidizing and Reducing Agents, Handbook of Reagents for Organic Synthesis , John Wiley and Sons, New York 1995.

[2] Caruthers W. Some Modern Methods of Organic Synthesis Cambridge University Press; 3rd edition, January 1987.

[3] Cooke M. Computer analyses of qualitative data: A literature review of current issues, Australian Journal of Advanced Nursing, 10(1), 10-13.

[4] Girratano, J. and Riley, G. *Expert Systems: Principles and Programming* PWS Publishing Co., Boston, Massachusetts, 1998.

[5] Gilbert J. and Martin S. Experimental Organic Chemistry (2nd Ed.)1998.

[6] Inxight Software – Unstructured Data Management Solutions. Available at http://www.inxight.com/ (visited May 5, 2003).

[7] Larock R., Comprehensive Organic Transformations: A Guide to Functional Group Preparations, John Wiley and Sons, New York 1999.

[8] Martin, B., Subramanian, G.H., and Yaverbaum, G.J. Benefits from Expert Systems: An Exploratory Investigation. *Expert Systems with Applications 11*, 1 (1996), 53-58.

[9] Martin, J. and Oxman, S. *Building Expert Systems*. Prentice Hall, Englewood Cliffs, New Jersey

[10] Morvant M, Halterman R, and Phillips M, Organic Chemistry Labaratory Manual.2002.

[11] Pasto D. and Johnson C. *Organic Structure Determination*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1989.

[12] Riley, Gary. CLIPS: A Tool for Building Expert Systems. Home page. Available from http://www.ghg.net/clips/CLIPS.html. (visited April 30, 2003).

[13] Shriner, Hermann C, Morrill T, Curtin D, and Fuson R. *The Systematic Identification of Organic Compounds*, 7th edition, John Wiley and Sons, New York 1998.

[14] Zubrick J. *The Organic Chemistry Lab Survival Manual*, (4th Ed.,) John Wiley and Sons, New York, 1997.

Copyright © 2004, American Society for Engineering Education

Biographical Information

Mario A Garcia finished his Ph.D. in Computer Science at Texas A&M University in August 1997. He was a professor of Electrical Engineering and Computer Science for 15 years at Tecnologico de Saltillo and Tecnologico de Monterrey (ITESM) in Mexico. He has been working as an Assistant Professor at Texas A&M University-Corpus Christi since 1999. His areas of interest are Artificial Intelligence, Software Engineering, and recently, Information Assurance.

Copyright © 2004, American Society for Engineering Education