Rule-Based Expert System to Determine the Likely Diseases in an Air-Polluted Environment

R. V. Díaz[1, 3], Zenaida García[2, 3], Mateo Lezcano[2, 3] and Elena Rosa Dominguez[2, 3]

[1]rvdg@nuclear.inin.mx
[2]{zgarcia, mlezcano, erosa}@uclv.edu.cu

[1]Instituto Nacional de Investigaciones Nucleares (ININ),
Carretera México Toluca S/N, La marquesa Ocoyoacac, México. C.P. 52750
Tel.: +52 53 29 72 00 ext 2646, fax: +52 53 29 73 32

[2]Universidad Central "Marta Abreu de las Villas"
Carretera a Camajuaní km 5.5. Santa Clara, Villa Clara, CP 54830
Tel. 53 42 281194/ 281630, Fax: 53 42 281608

[3]Instituto Tecnológico de Toluca (ITT)
Av. Instituto Tecnológico s/n ExRancho La Virgen, Metepec, México CP: 52140
Tel. 01 722 208 7 224 ext 340

Abstract. In this paper we present an expert system which relies on production rules and aims to determine the set of diseases which might appear in a polluted environment. The main contribution of this study lies in the quality of the acquired knowledge by using rules as the inference engine. Experimental studies comprised the application of PIXE (X-ray Emission induced by Particles) techniques. The expert system is intended to be used by chemical analysts with an environmental profile, yielding an accurate representation of the knowledge base and leading to several important conclusions.

1 Introduction

Among the several branches that Artificial Intelligence encompasses, the Expert Systems (ES) have climbed to a unique position owing to their successful applications to a wide variety of problems as well the speediness in getting the problem solution and a higher quality of the solutions obtained [1, 2, 3]. Such features turn the ES into an indispensable tool for decision making in, for instance, the field of risk management, wherein quick and reliable decisions are needed. Thus the ES guarantees its inclusion within the set of strategies for risk management. A clear problem statement is fundamental when studying the atmospheric pollution bearing in mind that the main goal is to avoid the possible health risks by the contaminant powers present in it such as the suspended particles with an aerodynamic diameter less than or equal to 2.5 micrometers. ($PM_{2.5}$).

© A. Gelbukh, Á. Kuri (Eds.)
Advances in Artificial Intelligence and Applications
Research in Computer Science 32, 2007, pp. 442–449

Received 16/06/07
Accepted 31/08/07
Final version 30/09/07

Considering that a great deal of countries, mostly located in Latin America, do require this type of research study, the envisioning and implementation of strategies tending to assure a right decision making is compelling, therefore achieving a better overall air quality and equipping the population with a fat better life style. In Mexico, relying on trustable data to decision making is of the utmost importance so as to enhance, for example, the development of Mexican standards of the air pollution as well as counting on sound foundations for the development and evolvement of programs such as the one of the Metropolitan Indicator of Air Pollution (IMECA, acronym in Spanish) that are suitably applicable without the need of copying models coming from other countries having a very different population in terms of customs and idiosyncrasy [5, 6]. The expert system created for such purposes allows determining the likely diseases which may arise due to the chemical elementary composition contained in the $PM_{2.5}$ and will facilitate the decision making process in such settings.

2 An Overview to the Environmental Problem Leading to the Expert System

The elementary chemical composition of the $PM_{2.5}$ is a potential danger for the population because of the inhalation risk. This was an essential requisite to be considered for building the expert system. This section focuses on describing the characteristics and methodology taken from the environmental domain.

2.1 Main Features

a) The sensitive population group was represented by children from 2 to 6 years old as well as those ranging from 6 – 12 and adults with an average age of 70 years old.
b) The sensitive groups endure, on a daily basis, contamination episodes mostly provoked by the $PM_{2.5}$ being present in their dwelling place [7, 8, 9].
c) Pollution sources: The exposition origin to these substances is of natural and anthropogenic genesis (lots of industries, automobiles, etc.)
d) Exposition Point: The atmospheric air.
e) Conveying Mechanism of Contaminant Powers: The air.
f) Exposition Ways: The inhalation of the polluted air.
g) Period of Observation: Dry, cold season.

2.2 Methodology

a) A place of interest where the contaminant $PM_{2.5}$ substances reside may damage the sensitive population groups.
b) Samples of the contaminant powers were picked up and subsequently analyzed by considering the Protons-Induced X-Ray Emission (PIXE) technique. The achieved

elementary chemical composition is: S, Cl, K, Ca, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, Br and Pb [8, 9].

c) The inflicted population risk by the inhalation of this contaminant was determined after regarding the Exposition Dose (ED) and the Reference Doses (RD) according to [10, 11]. The expression governing this step is the ensuing:

Risk (R) = (DE) (mg/kg/day) / DR (mg/kg/day)

The outcomes for the three sensitive groups are pictured in Table 1. From the obtained elemental risk, the likely diseases that may be caused by an element either alone or in conjunction with others hitting the sensitive groups are calculated. In total, 79 relationships were spotted. Some of them are illustrated in Table 2. After arranging the knowledge expert, the Expert System was built by using production rules as the Knowledge Representation Form (KRF).

		RISK (R) = DE / DR	
ELEMENT	ADULTS	6-12 year-old Children	2-6 year-old Children
		DS (mg/kg/day)	
SULFUR	0,03	0,09	0,09
CHLORINE	0,22	0,69	0,68
TITANIUM	0,16	0,16	0,16
VANADIUM	7.84E-04	2.45E-03	2.41E-03
CHROMIUM	0,02	0,05	0,05
MANGANESE	0,04	0,13	0,13
NICKEL	3,66E-05	1,14E-04	1,13E-04
COPPER	1.33E-04	1.33E-04	1.33E-04
ZINC	1,50E-05	4,70E-05	4,63E-05
LEAD	0,68	0.68	0.68
Overall Exposition Index	1,15	1,81	1,79

Table 1. Risk representation for the three sensitive groups.

3 Results and Discussions

The aforementioned methodology gave rise to the expert system which aims to establish a relationship between elemental chemical composition contained in the suspended particles $PM_{2.5}$ and the several diseases which those may bring about. The Rule-Based System (RBS) is made up of 480 rules from which 243 of them are used in the diagnosis of the possible diseases and the rest allow the functioning logic to perform well. The Inference Engine (IE) was supplied by the WUCShell system version 1.0 [15]. Figure 1 shows the functional structure of a Rule-Based Expert System using WUCShell.

An illustration of the rules section of the ES is presented as follows:

```
Rules
    Rule 7
      If element = 'S-Ni'
      Then
        SNi_Adult := 17.1*concentrationS+61*concentrationNi
    End;
    Rule 8
      If SNi_Adult > 1
      Then
        Disease: = 'asthma'
```

DISEASE	ELEMENTS [10, 12, 13, 14]
Asthma	S, Ni
Hormonal metabolism alterations	S
Bronchopneumonia	S
Cardiac damages	S, Cl, Ni
Damages on liver functions	S
Damages on kidney functions	S, Pb
Headaches.	S; Cl, Mn, Cu
Thoracic pain and obstruction.	S, Cl,
Bronchitis	S, V, Mn, Cl, Ni
Fever	S
Nose irritation	S, V, Cr
Throat irritation	S, Cl, V
Respiratory seizures	S, Ni
Death	S, Cl, Cr, Pb
Itchiness or burning sense of nose and throat	Cl

Table 2. Disease-elements relationships.

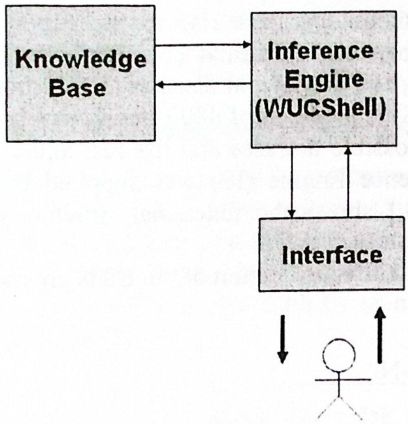

Figure 1. Functional structure of a Rule-Based Expert System using WUCShell.

Rule 7 expresses the calculation of the sum of the elemental concentrations for sulfur (S) and nickel (Ni). Rule 8 compares the sum of the risks with one (1). If the risk exceeds one due to the sum of S + Ni, then the chances are that the adults come down with asthma because of the inhalation of S and Ni. That possibility is null the other way around.

The ES attributes exhibited in the rules conclusions are inferables, i.e. they have to be checked across the inference engine. Other attributes are asked by the user and that's why the ES will prompt several times for the specification of the proper values as the inference process goes on. An excerpt of the latter type of attributes can be observed in the following lines.

```
Asks
  ASK element:
    'Por favor seleccione uno de los compuestos que se le muestra.'
    Domain
      'S', 'S-Ni', 'S-CL-Ni', 'S-CL', 'S-Pb', 'V', 'S-CL-Mn-Cu',
      'S-V-Mn-CL-Ni', 'S-Mn-Ni', 'S-V-Cr', 'S-CL-V', 'S-CL-Cr-Pb',
      'CL', 'CL-Ti-V', 'Ti', 'Cr', 'Cu', 'Mn', 'Ni', 'Cu-Zn', 'Pb'
  ASK concentrationS:
    'Especifique la concentración química del azufre (mg/m3):'
    Because
      'Con el uso de este sistema usted puede determinar el riesgo /n',
      'que enfrenta la población al respirar diferentes tipos de    /n',
      'contaminantes.                                               /n',
      'Debe especificar la concentración química del azufre para   /n',
      'el cálculo del índice de riesgo./n'
  ASK concentrationNi:
'Especifique la concentración química del níquel (mg/m3)'
```

The WUCShell inference engine is an objective-driven search tool by using the depth-first search method. The action block of the knowledge base is displayed next.

Actions
Find disease
End.

This is the starting point where the inference process begins. Notice that the sole objective to be inferred is the disease. A search process is launched by the inference engine which looks for that target alone, finishing only when a value is assigned to such attribute (successful test) or arriving to a state wherein no more rules are explorer (failure test). Some of the windows of the user interface of the WUCShell tool involved in the execution of the ES for a sample where the user picks up both the sulfur and nickel components (S-Ni) are portrayed in Figure 2.

When considering the multiple choices presented in Table 2 for the three sensitive population groups, it was inferred that none of the cases residing in the knowledge base are at stake of enduring any of such diseases. The only combination for which the likelihood of enduring any disease for population groups 2 and 3 (children-related) was that of (S, Cl, Cr, Pb). The disease in question is death. The same combination of elements is harmless for adults. Looking at the results bore by the ES when regarding a single element, it is expected that there is no chance of suffering any disease since the elementary risk is low. However the synergism of various elements triggers a risk increase which, in accordance with the literature, the risk is higher than the sum of the individual risks. The actual risk value owing to the synergic properties of the elements was not regarded in this study. The disease caused by the conjunction of elements S, Cl, Cr and Pb yield a risk having a sum greater than one and hence the aforementioned comment holds. With the aim of testing the ES for another study areas and considering elementary concentrations of the same contaminant (PM2.5), you can bear the results shown at Table 3.

DISEASE	ELEMENTS	LOCATION	LIKELIHOOD	REF.
Bronchitis	S, V, Mn, Cl	Ticomán (ZMCM)	NO	[16]
Bronchitis	S, V, Mn, Cl	Monterrey N.L. Mexico	SI	[17]
Kidney damage	S, Pb	Monterrey N.L. Mexico	SI	[17]

Table 3. Relation between the possibility of enduring diseases and several locations of the Mexican Republic (adults only)
ZMCM: Metropolitan Zone of Mexico City (acronym in Spanish). N. L. Nuevo León

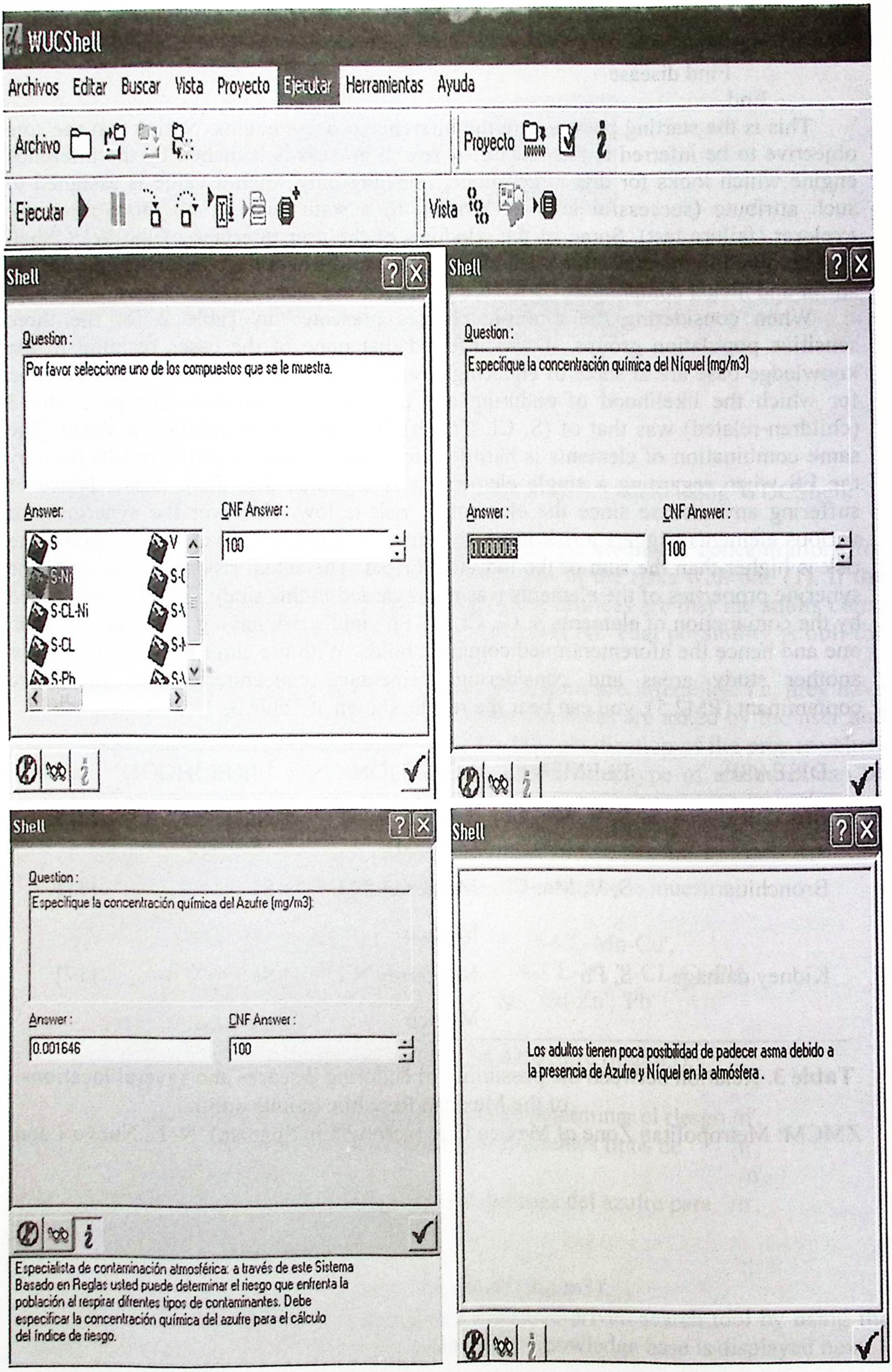

Figure 2. Different windows making up the execution of the ES for a sample.

4 Conclusions

1) The WUCShell tool allowed building the expert system in a fast and straightforward fashion.
2) Since the developed ES relies on sound data gathered by using reliable scientific methodologies, its own results are reliable as well.
3) Using this ES enables to bear quicker and more accessible results in such a way that it can be considered as an important tool for decision-making, i.e. an strategy to the risk management brought about by the inhalation of the $PM_{2.5}$.
4) The ES allows knowing the health risk (likely diseases the population may endure) as well as the emission sources coming from the chemical elementary composition. This was conducted in a way that allows exercising some degree of influence in the control or removal of the sources issuing this contaminant to the atmosphere.

References

1. Hatzilygeroudis I., (2004): An Expert System with Certainty Factors for Predicting Student Success. Lectures Notes on AI, Springer-Verlag.
2. Hayes, P., The Naïve Physics Manifesto. in: D. Michie (Ed.), Expert Systems in the Microelectronic Age. pp. 242-270, Edinburgh University Press, Edinburgh
3. Peña C., (2001): Toxicología Ambiental, Evaluación de Riesgos y Restauración Ambiental., The University of Arizona, 1996-2001
4. Mackenzie L., (2005): Ingeniería y Ciencias Ambientales, Mc. Graw-Hill Interamericana.
5. www.sma.gob.com
6. www.ine.com.mx
7. Cicero Sabido R., (1996): Contaminación Atmosférica y Salud en la Ciudad de México / Riesgos Ambientales Para la Salud en La Ciudad de México. Programa Universitario del Medio Ambiente, UNAM, p: 466.
8. Aldape F., (1991): Two Year Study of Element Composition of Atmospheric Aerosols in Mexico City. International Journal of Pixe, Vol.1, No.4 (1991)373-388.
9. Aldape F., (1993): Temporal Variations in Elemental Concentrations of Atmospheric Aerosols in Mexico City. NIM B75 304-307.
10. http://www.atsdr.cdc.gov/hac/PHA/isladeviquesspan/cdb toc ,html. ATSDR, Agency for Toxic Substances and Disease Registry, 2003. Evaluación de la vía del suelo / Campo de Bombardeo de la Isla de Vieques. Vieques Puerto, Febrero 7.
11. CEPIS/OPS, (2001): Curso de Auto instrucción de Evaluación del Riesgo Asociado a Contaminación del Aire.
12. http://www.lenntech.com.
13. EPA, U.S, (1991): Environmental Protection Agency. Risk Assessment for Toxic Air Pollutants: A Citizen's Guide - EPA 450/3-90-024, March.
14. www.toxnet.com. TOXNET, Toxicology Data Network.
15. Lezcano Brito, M. (1999): Manual para el uso de la máquina de inferencia WUCShell, Versión 1.0. Universidad Central de Las Villas, Santa Clara, Cuba.