

Discovering epistemological axes for academic programs in computer science through network analysis

Luis Casillas Santillán
Computer Science Department
University of Guadalajara, Mexico
luis.casillas@red.cucei.udg.mx

Alonso Castillo Pérez
Computer Science and Electronics Division
University of Guadalajara, Mexico
divec@cucei.udg.mx

Abstract: Academic programs in higher education institutions (HEI) are constantly under construction. They are never finished at all. As reader could agree, the current merging of knowledge in present societies is inexorable. The available cyber-infrastructure is able to support new approaches in science development. Hence, academic programs in HEI are clearly under the pressure to be constantly modernized in order to provide valid and up to date knowledge. Every academic program has this challenge in present societies. Our approach is aimed at providing a practical method for discovering the epistemological axes through the use of some notions from complex networks analysis. We worked on “Computer Science” programs in the University Center for Exact Sciences and Engineering from the University of Guadalajara. Nevertheless, the proposed method could be easily extrapolated to different programs in HEI, due to similar nature of knowledge in diverse knowledge branches.

Keywords: Complex networks, knowledge discovery, knowledge management, epistemological axes, clustering coefficient.

Descubriendo ejes epistémicos para programas académicos en Ciencias de la Computación a través del análisis de redes

Resumen: Los programas académicos en las instituciones de educación superior (IES) están constantemente en construcción. Nunca están terminados del todo. Como el lector podrá estar de acuerdo, la actual fusión de conocimiento en las sociedades presentes es inexorable. La ciberinfraestructura disponible es capaz de apoyar los modernos enfoques en el desarrollo de la ciencia. Por lo tanto, los programas académicos en instituciones de educación superior están claramente bajo la presión de ser constantemente modernizados para ofrecer conocimientos válidos y actualizados. En las sociedades actuales, todo programa académico tiene este problema. Nuestro enfoque está dirigido a proporcionar un método práctico para el descubrimiento de los ejes epistemológicos a través de la utilización de algunas nociones propias del análisis de redes complejas. Hemos trabajado en programas de "Ciencia Computacional" en el Centro Universitario de Ciencias Exactas e Ingenierías de la Universidad de Guadalajara. No obstante, el método propuesto puede ser fácilmente extrapolado a diferentes programas en IES, debido a la naturaleza consistente del conocimiento en las diferentes ramas de éste.

Palabras clave: Redes complejas, descubrimiento de conocimiento, gestión del conocimiento, ejes epistemológicos, coeficiente de agrupamiento.

1. Introduction

Most of the activities developed by humans are based on sets of concepts and the relationships among such knowledge elements. The human's understanding of the environment is defined by the concepts and ideas acquired before. These elements are organized as nets of perceptions in which the relationships are modeled by links of proximity.

Since 1960's, computers had been used formally in the scientific field. The tasks that used to be assigned to computers were mainly connected to calculations in iterative, recurrent and concurrent contexts. More recently, the advances in computers' graphics allowed improved simulations. Nowadays, computers are not the actor using "brute force" anymore. Of course, they will do the calculations needed; however, the main goals have changed. A goal for computers in this context is to support the collaboration among scientists, in the already mentioned fields for e-science (Casillas and Daradoumis, 2012).

The report from Atkins et al. (2003) is highly focused in the exploitation of the available cyber-infrastructure. According to them, there is a revolution in science due to advances in information and communication technologies. This revolution is based in the innovative capabilities to successfully **emulate reality in the digital dimension**. Specifically, these authors argue *"...the classic two approaches to scientific research, theoretical / analytical and experimental / observational, have been extended to in silico simulation and modeling to explore new possibilities and to achieve new precision..."*; with important achievements in Forestry, Ocean Science, Environmental Science and Engineering, Space Weather, Computer Science and Engineering, Information Science and Digital Libraries, Biology / Bioinformatics, Medicine, Physics, Astronomy, Engineering, Materials Science & Engineering, and Social & Behavioral Sciences.

For Jankowski (2007) e-science is strongly supported by key aspects of the information and communication technologies, which are:

1. *"International collaboration among researchers;*
2. *Increasing use of high-speed interconnected computers, applying Grid architecture;*
3. *Visualization of data;*
4. *Development of Internet-based tools and procedures;*
5. *Construction of virtual organizational structures for conducting research;*
6. *Electronic distribution and publication of findings "*

Current scientific context, technological advances, and the socialization of information technologies and communications (social nets, blogs, video repositories, augmented reality, etc.); are the core of new understandings in the development of modern academic programs in current and future HEI. Our proposal is aimed at discovering the knowledge nucleus supporting an academic program, using practical and / or automated steps. These nuclei represent the epistemic axes for the academic programs in HEI. According to the arguments of Popolo (2012) and Martin (2003), an epistemic axe is a specific context and a precise trajectory in a Bayesian approach for reality. This perspective agrees with our approach for academic nuclei.

In an effort to discover the body of knowledge that forms the training core for "Computer Science" programs from the Department of Computer Sciences of the University of Guadalajara, we had made a study aimed to discover the essential branches of education for these programs, as well as their cognitive infrastructure.

Using techniques for discovery and extraction of knowledge stored in humans, a group of professors were invited to collaborate in this study. The results produced at the time, are certainly interesting.

2. Strategy

The first step was to identify professors and specialists who would be collaborators in the study. We agreed that they should be individuals possessing a deep understanding of the syllabus studied, both in the pedagogical dimension as in research or professional areas. Reaching an “elite” of approximately 10% of faculty and researchers of the program studied. See figure 1. Because this study could be conducted through electronic means (email, forums, etc.); collaborators may be in remote locations and continue to participate throughout an internet access.

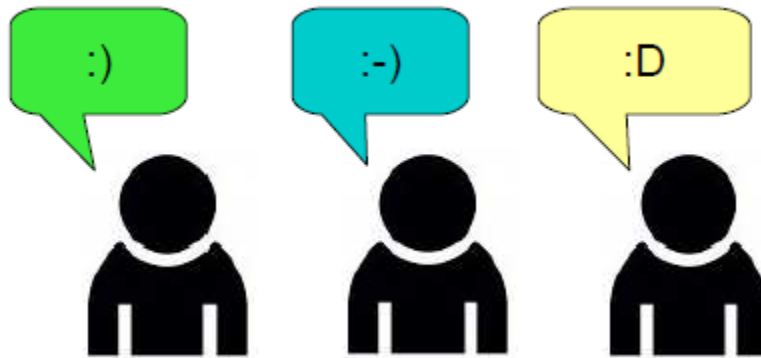


Figure 1. A group of distinguish professors is invite to collaborate in the study. Most of the invited professor gladly accepted. Professor rejecting the invitation excused arguing they had no time.

Collaborators were asked to identify the main areas of knowledge related to the program studied. Preferably a list of 7 ± 2 items (Miller, 1956), containing only the name of each branch. These knowledge branches did not have to be similar to the current subjects in the program. Now they must look at the curriculum as a critical mass of knowledge that would be fragmented. At this point, they were requested the main branches. As shown in figure 2.

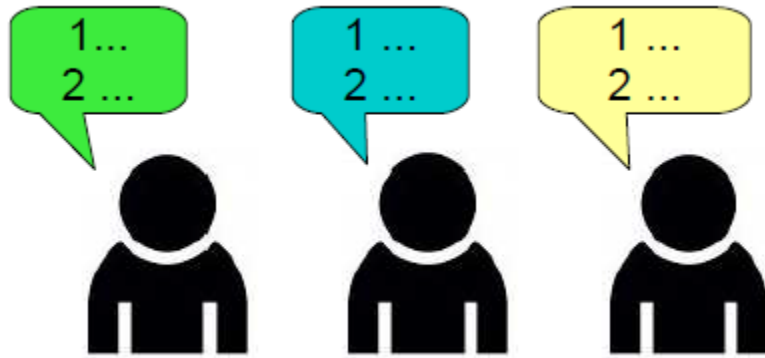


Figure 2. Professor collaborating produced the list of knowledge items for the academic program in which he/she was participating.

The following step was to collect the names of branches defined by each collaborator. These names had some naming difference and they were merged, combined, unified and refined to a common nomenclature that meets all of the collaborators proposals. As sketched in figure 3. The weight assigned to contributions was democratic: the knowledge engineer viewed only the names of the branches without having access to the collaborator's name. The final product of this step is a list of 7 ± 2 elements, which represents a common ideological kernel from collaborators.

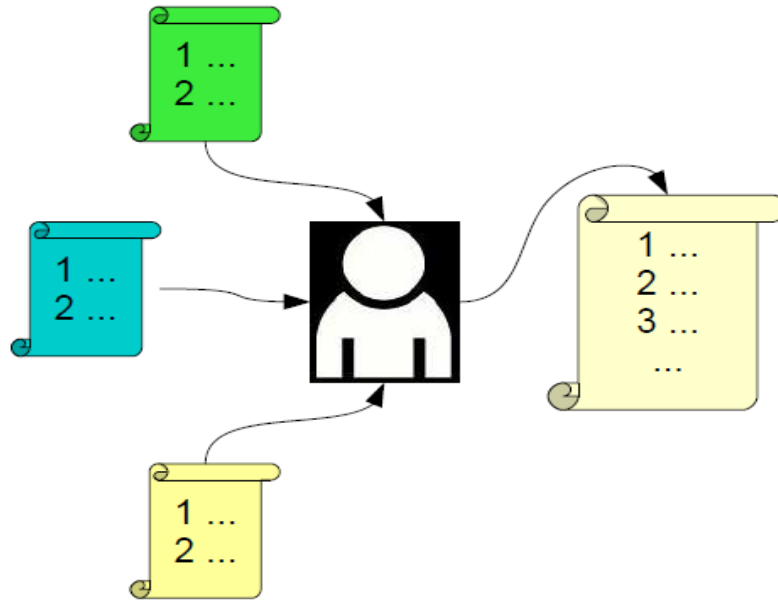


Figure 3. Knowledge engineer proceeds to collect the lists of knowledge items (branches) produced by professors. The challenge for this engineer at this point, is to produce a compact list (7 ± 2 items) which merges the branches proposed. Discovering those are common, similar or opposite each other, without missing the valuable awareness provided by collaborators.

The next activity was to distribute, among collaborators, the consolidated list of branches. As shown in figure 4. Once they were distributed, collaborators had to divide each of branches from unified list into sub-branches or milestones of knowledge. The number of sub-branches should again be around 7 ± 2 elements. So far the collaborators have made a classical decomposition of knowledge: a hierarchical approach. This was sketched in figure 5. Humans make this kind of decomposition almost inherently.

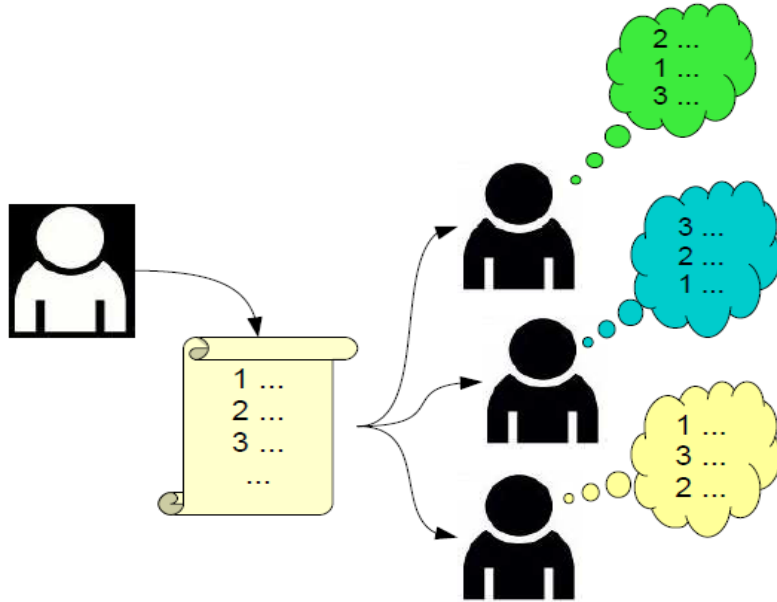


Figure 4. Knowledge engineer transmits the merged list of items (knowledge branches) to collaborating professors. They gather these elements and proceed to the next stage of the study.

The following step is to re-make a major effort to combine, unify and refine the milestones (subbranches) that have been provided by each collaborator. The final product of this effort is a list of 7 ± 2 milestones per branch. Again, this represents an ideological kernel of contributors, but now for the milestones.

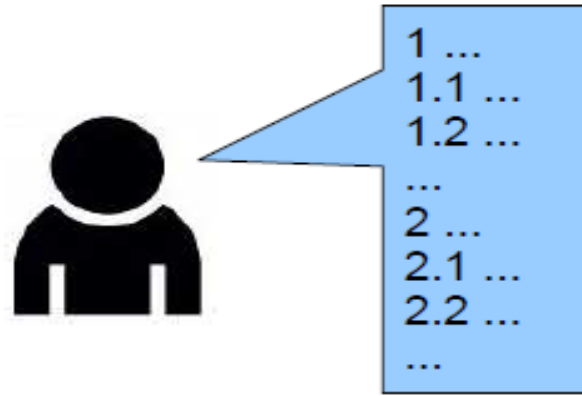


Figure 5. At this point, every collaborating professor produce coherent sub-branches for each branch received. These sub-branches are the milestones he/she considers as the atomic concepts composing every knowledge branch.

The unified list of milestones is sent to collaborators to perform their last task: find dependencies and / or influences among milestones. This whole process is modeled in figure 6. Every collaborator can set all the bounds he/she considers as relevant, regardless of the branch to which milestones belong. As long as they are rational or consistent links, but this should not be cause of concern if collaborators are characters with solid experience. As shown in figure 7.

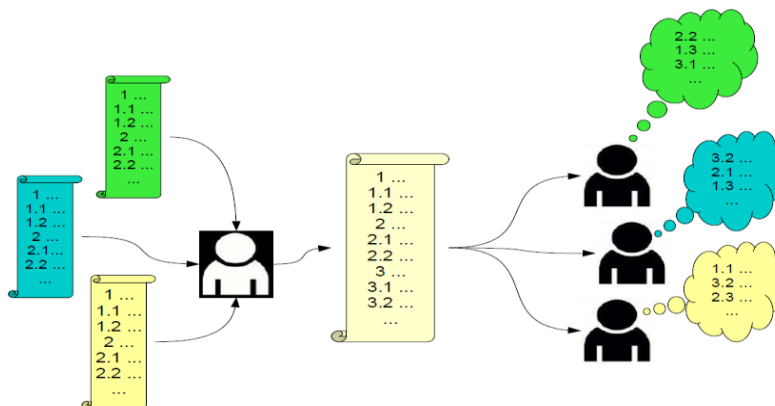


Figure 6. Once again knowledge engineer performs a collection of lists coming from collaborating professors. These lists contain the knowledge milestones they considered as sub-branches. Now, knowledge engineer proceeds to

merge, unify and conjugate all the milestones produced. This challenge was supported by electronic worksheets and their formulae, filters and macros. Resulting merged list is now resend to collaborating professors in order to start their last task.

Once collected the information from relationships, we proceeded to build a directed graph containing all the information. The nodes of this graph are the milestones, and the edges represent the relationships of influence or dependence among them. *The understanding of the problem has ceased to be hierarchical to become a networked approach.*

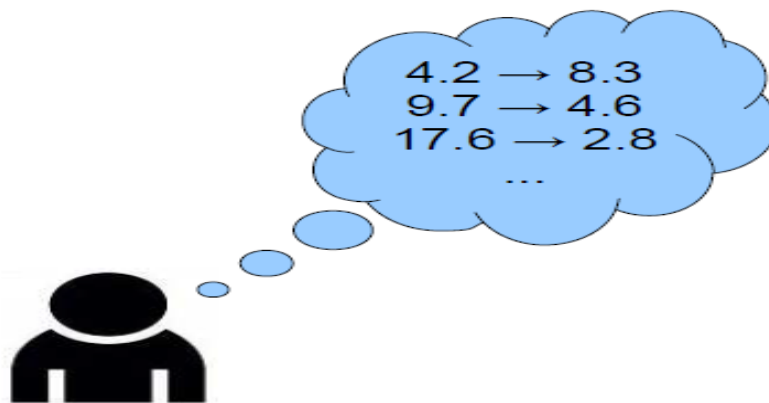


Figure 7. Every collaborating professor establishes influences and dependencies among the knowledge milestones provided in unified list. These binary relations would be combined into a complex network structure. All the asserts provided by collaborators are used in the network.

Networked construction (see figure 8) enables the search of training cores from the academic program studied. The output degrees (the number of arcs leaving each node) indicate the influence of a milestone on others. The input degrees (the number of arcs entering each node) indicate the dependence from other milestones. The most influential milestones, as well as those less dependent

are, on a preliminary review, the main aspects shaping the training core of the program. Figure 8 shows the network built from knowledge pieces discovered in this process, which guided us to find out the training cores for programs in computer science.

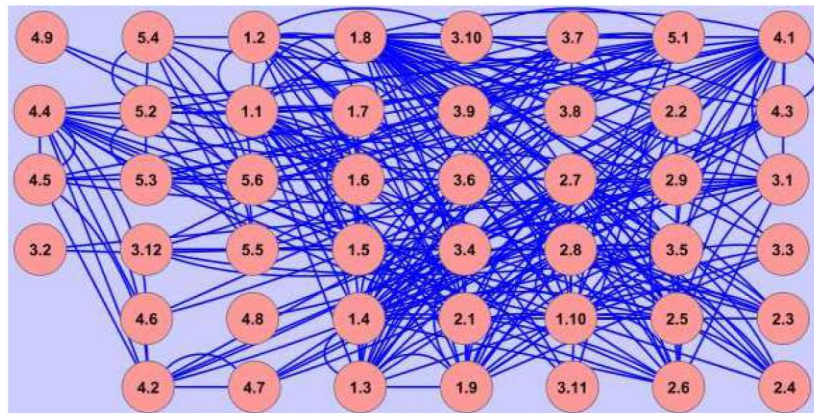


Figure 8. Network built from binding and linking the milestones and relations provided by collaborators.

Input and output degrees may incidentally acquire magnitudes disconnected from the rules that explain the studied reality. In fact, this network tends to match the form of a scale-free network (Barabasi, 2003). Thus, we complemented these results with the discovery of hub nodes in the graph. By discovering hub nodes, we find out relationships among families of nodes in the neighborhood which share access to common nodes. This neighborhood is called cluster. The clustering coefficient (Watts and Strogatz, 1998), is calculated locally (for each node) and globally the average of all the coefficients calculated for each node locally).

Based on the arguments from Watts and Strogatz (1998), the formula for clustering coefficient in directed graphs is: $C_n = \frac{p_n}{(v_n (v_n - 1))}$, and the formula for non directed graphs is $C_n = \frac{p_n}{v_n}$. Where C_n is the clustering coefficient for a node n , p is the number of pairs formed between the node's neighbors

observed, v_n is the number of neighbors of node observed. Figure 9 sketches a simple example for non directed graphs.

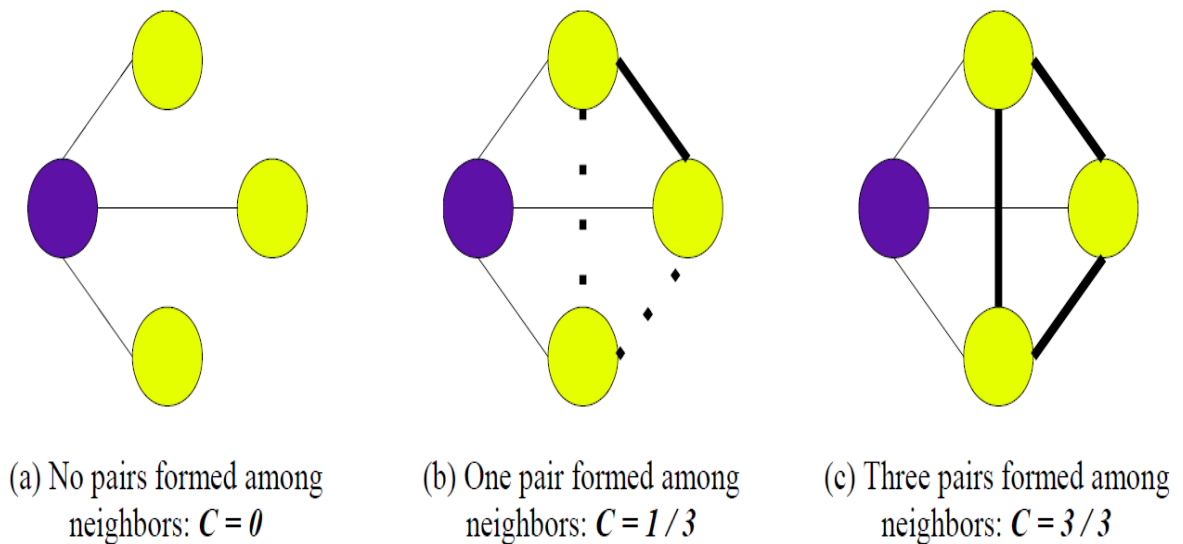


Figure 9. Clustering coefficient for a non directed graph.

The clustering coefficient is measured within the interval $[0, 1] \in \mathbf{R}$ and indicates the strength of the group that maintains a common node at the center. The local coefficient speaks of the capacity of a specific node to be a hub node; the global ratio refers to the cohesion of the entire structure. A good level in the overall coefficient (0.51 to 1) and the discovery of sub-networks involving hub nodes, lead to the threshold of *training cores*.

The networked stages of this strategy were supported by software made by us, specifically for these tasks.

The knowledge engineer must compare the information about the influences and dependencies between milestones and the awareness regarding hub nodes to form groups among them. This process should be carried out among the knowledge engineer and experts.

3. Results

Using this strategy, they were identified the *epistemological axes* for academic programs in computer science from our department:

Computer Sciences

Networks, Communications and Distributed Systems

Intelligent Systems

Information Systems (for bachelor's degree in informatics)

Systems Programming (for bachelor's degree in Computer's Engineering)

Technology Management

These results have enabled the development of new approaches for our academic program in computer science and are the base for developing global examinations for computer science graduates.

Diverse professors specialized in the area have reviewed our results and manifest highly acceptance to this proposal of academic axes.

4. Conclusions and future work

This challenge was undertaken in order to solve specific necessities in our academic context, but we were able to realize that this method could be easily transported to new academic scopes and allow diverse academic programs to undertake the very same strategy. We keep on discovering knowledge elements laying underneath these networks. For instance, the nodes with higher output degrees should be sent to first stages of formation and those with higher input degrees should be sent to last stages in the program.

The results we reached, were highly influential in the development of new proposal for significant content renewal for the bachelor's degrees in the Electronics and Computer Science Division from the University of Guadalajara.

The novel approaches for these academic programs will soon be started in our institution.

We believe that many of the stages could be automated. In order to finish these work, some human supervising and adjustments had to be made. Nevertheless, many tasks are clearly assignable to machines. That would be the following stages in the process we are undertaking.

Acknowledgements

Authors would like to thank professors from Computer Science Department for their participation in this study, as well as the coordinators of the academic programs involved.

References

Atkins, D.; Droegemeier, K.; Feldman, S.; Garcia-Molina, H.; Klein, M. L. and Messina, P.(2003). "Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure". Washington, DC: National Science Foundation.[on line] Retrieved June 16th 2011, from www.nsf.gov/cise/sci/reports/atkins.pdf

Bales, R. F. (1970). *Personality and interpersonal behavior*. New York: Holt.

Casillas, L. and Daradoumis, T. (2012) "An ontology structure for gathering and sharing knowledge among. scientists" *Collaborative and Distributed-Research: Innovations in Technologies, Strategies and Applications*. IGI Global Publishing, Hershey, Pennsylvania, USA.

Jankowski, N. (2007) "Exploring e-science: An introduction. *Journal of Computer-Mediated Communication*" 12(2), article 10. [on line] Retrieved June 15th 2011, from <http://jcmc.indiana.edu/vol12/issue2/jankowski.html>

Martin, K. (2003) "Epistemic Motion In Quantum Searching". Oxford University Computing Laboratory.

Miller, G. A. (1956) "The Magical Number Seven, Plus or Minus Two. Some Limits on Our Capacity for Processing Information". American Psychological Association. *Psychological Review*, Vol. 101.

Popolo, D. (2012) "A New Science of International Relations". Ashgate Publishing.

Watts, D. and Strogatz, S. (1998) "Collective dynamics of 'small-world' networks". *Nature* 393 (6684) pp 440–442.

Biographical notes:

Luis Casillas Santillán holds a Ph.D. and a master's degree in Information and Knowledge Societies from the Open University of Catalonia, Spain; as well as a master's degree in Information Systems and a B.Sc. in Informatics from the University of Guadalajara (UG), Mexico. He has been working as associate professor for 17 years in the Computer Science Department from the UG and published various papers and scientific chapters in diverse journals and books. He serves as member of the editorial board for the *Journal of Knowledge Engineering: EXSY* and as reviewer for diverse journals and other scientific publications in knowledge engineering, computers science and ICT in education. His research interests are: knowledge gathering and representation, bio-inspired systems, expert systems, complex networks analysis, and soft-computing.

Alonso Castillo Pérez holds a master's degree and a B.Sc. in Electrical Engineering from the University of Guadalajara (UG), Mexico. He has been professor in diverse levels and programs, from basic education up to graduate programs in structures, industrial systems, potency systems and the doctorate program in engineering and technology in the UG. He worked for the Federal Commission for Electricity (CFE). He has held several positions in the UG and the CFE. He authored plenty of books, about mathematics and electrical engineering. Besides, he wrote a novel about the Mexican rural school. He is currently Director for the Electronics and Computer Science Division from the University Campus for Exact Sciences and Engineering of the UG.



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 2.5 México.