22

Different Approach to Information Technology - Teaching the Intelligent Systems Course

Ana Madevska Bogdanova and Nevena Ackovska
University Sts Cyril and Methodius,
Faculty of Natural sciences and Mathematics
Institute of Informatics
Macedonia

1. Introduction

The curriculum of Computer Science module at our institute attracts students with broader mathematical background, who are eager to learn something more and different. This module consists mainly of subjects that are concerned with problem solving (mathematical, algorithmic), with the aid of procedural and Object-Oriented Programming paradigms. However, there are problems that can't be solved by the aforementioned techniques. When we confront the students with a problem of pattern recognition or prediction represented by a stack of data with no mathematical or statistical model describing it, they realize that 'procedural' thinking will lead to no success. They may try to do some statistical modeling using common statistical analysis, but when they compare those results with the results obtained with a suitable learning method, they understand that the solution is not always in a set of instructions (procedures or classes and objects). So, a better approach is to use a 'learning machine' and see how it works on a particular problem.

The real life problems usually are solved by finding a good answer when the perfect answer is not possible to obtain. This line of thinking was not offered by the undergraduate courses in the Institute of Informatics before 2004. This gave us the idea to design the Intelligent Systems course, which should enable students to receive another viewpoint for solving complex problems.

To explain how we managed to organize this particular course, we structured this chapter in 6 parts, including this introductory part. We begin our elaboration with the course description in part 2, where we describe the course organization and its key components in particular. In part 3 we elaborate the methodology used in this course. In part 4 we refer to the outcomes and in part 5 we describe future development of the Intelligent Systems course. The conclusions are given in part 6.

2. Course description

The Intelligent Systems (IS) course is organized in two parts: theoretical part with 2 classes per week and exercise part with 3 classes per week. It is carried out in an overview fashion,

where 4 main parts are introduced: Machine Learning (Pattern Recognition), Data Mining, Bioinformatics and Robotics. Every year we explore the chief up-to-date developments in the above mentioned topics. Every new generation deals with a new set of exercises, i.e. the new hot problems (Madevska-Bogdanova & Ackovska, 2008).
The teaching material is recruited from different sources – textbooks, journal papers, conference papers, web pages with high level of credibility etc. Textbooks are used in understanding the theoretical background of the offered topics (mathematics, statistics, Bayesian models, numerical optimization,…). Journals and conference materials are considered when explaining the latest achievements in pattern recognition in Bioinformatics (gene, promotor recognition), Data Mining (discovering new knowledge in given data sets) and Robotics. We also use open source software support. For the Machine Learning techniques we use SVM light (Thorsten, 2003) and Weka (The University of Waikato, 2004) for Data Mining. Different web sources are used for gathering material from real life pattern recognition problems or different benchmark sets. We also exploit data set materials from ongoing or past projects from the European or USA Institutions that we collaborate with.
The course is designed to involve the students in the process of building systems for solving different kinds of problems. In the exercise part of the IS course each student has his/her own copy of the simulator on the laboratory PCs. The students go through the data set on their own, or in smaller groups, depending of the magnitude of the problem. Some of the exercises contain programming tasks about solving problems of the different topics.
In the sequel we shall elaborate the main constituent parts of the IS course.

2.1 Machine Learning

The first part of the course is about the Machine Learning (ML) techniques. Machine Learning usually refers to the changes in systems that perform tasks taken from Artificial Intelligence. We teach our students that the machine can 'learn' like the living creatures do – when the environment changes its structure or input data. Such tasks involve recognition, diagnosis, planning, robot control, and/or prediction. Computer Science students, with their solid mathematical background, for the first time in this module understand the potential of the mathematical disciplines such as numeric optimization, statistics, vector spaces, etc when used for solving practical problems.
We teach ML techniques mainly for dealing with the pattern recognition problems. We use Artificial Neural Networks (Bishop, 2006) and Support Vector Machines (Burges, 1998). It seems to be enlightening for the students to understand the other, non-standard ways of dealing with pattern recognition issues. They are confronted with problems from science, medicine, linguistics, trading, etc. They are taught to observe the similarities spread across this diverse set of problems (representation of the data). On the other hand, they are directed to understand the differences between the problems that lead to different ways of dealing with those problems (adjusting the parameters in the ML techniques). The performances of the given tasks in the above mentioned areas using ML techniques are much higher, compared to the standard statistical procedures used over the same problems. ML techniques are also good starting points for dealing with the Data Mining part introduced later in the course.
There are two good reasons to use the ML apparatus: we can simulate how humans learn, but there is also the engineering viewpoint concerning this issue - how to deal with a huge

amount of data in order to obtain new information when we don't have any mathematical model for creating a system that predicts the future behavior of this set of data.

Practical part. In order to practically illustrate the ML techniques for the classification problem, we use a simulator for a SVM classificator. Each student is given a data set for a certain problem (recognition of promoters in Arabidopsis Thaliana, recognition of mitochondrial sequences, classification of urinary calculi, ovarian cancer problem). The data sets of the mentioned problems are real data, derived from real life problems.

The students are divided in groups of two, approaching one of the problems with different SVM parameters (Kernel function and the appropriate parameters). The data are well prepared, preprocessed, so they can accomplish the task by the end of the class. Each group presents its results and the best results are discussed. The students draw some conclusions about the connection of the chosen parameters with the problem.

2.2 Bioinformatics

The second part of the Intelligent System course is Bioinformatics. Our students who possess sound mathematical background can find a great challenge in using mathematics and informatics tools in discovering meaningful sequences in the genetic material (Madevska-Bogdanova & Nikolik, 1999), (Madevska-Bogdanova et al., 2003). There are many ways to approach the problem of understanding the processes in the biological cell. For the purposes of the Intelligent Systems course, we are mainly concerned with the basic string processing aspects of DNA and RNA. The modeling of the genetics processes is done using the linguistic approach (Bozinovski et al., 2002).

Defining the terms bioinformatics and computational biology is not necessarily an easy task. Computational biology and bioinformatics are multidisciplinary fields, involving researchers from different areas of specialty. In order to be a good bioinformatician, it is important for the students to understand the terminology and basic processes behind the biological problems (Brown, 2002).

Many interesting problems arise out of sequence analysis. There are two different types of biological sequences studied in this class: DNA/RNA and amino acids (proteins) (Lewin, 2000), (Lodish, 2000).

It is important for the students to understand that by stringing together a simple alphabet of four characters together we can get enough information to create a complex organism (Ackovska & Madevska-Bogdanova, 2005)!

Practical part. Since the linguistic viewpoint of genetics processes is very natural for computer analysis, we use it for modeling in the student projects. The students are supposed to solve three types of problems. The first problem is to build a program module that should simulate the biosynthesis of proteins. The whole process has a DNA string as an input, and should obtain a protein as an output. As Figure 1 shows, the process is modeled by two pipelined Turing machines (Bozinovski et al., 2000), (Bozinovski et al., 2001). The first one is running the process of transcription: given an input DNA tape, produce an output mRNA tape. The second Turing machine is running the process of translation: given an input mRNA tape, obtain the polypeptide.

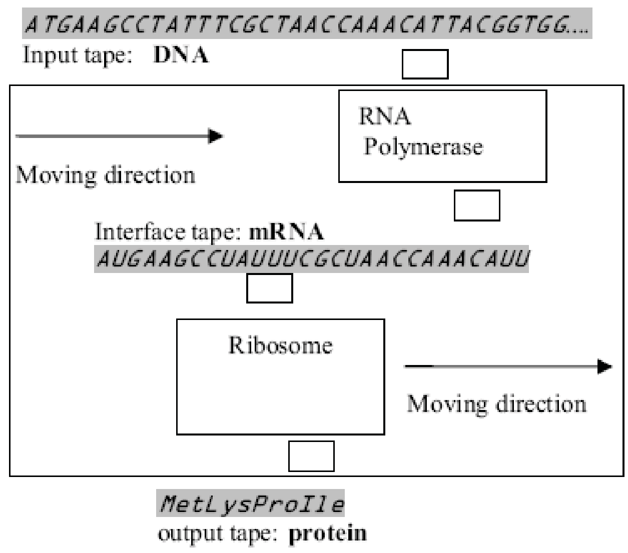

Fig. 1. Biosynthesis of proteins as a system of pipelined Turing machines

The solution of this problem could be written in any programming language. The students mostly give solutions written in Java or C++. Most of the student solutions contain 5 modules:

1. **check validity of DNA file: the** DNA file is consisted of 4 letters: A, C, T and G. It must contain the starting triplet ATG, from where the transcription begins.
2. **obtain mRNA file:** transcript from DNA alphabet, to RNA alphabet
3. **check validity of mRNA file:** mRNA files contain 4 different types of letters A, C, U and G. They should contain the starting triplet AUG. They must have a STOP triplet (UAA, UAG or UGA), to announce the end of translation. The number of letters between the starting and the ending triplets must be divisible by 3.
4. **obtain protein file:** translate from polynucleotide language (mRNA alphabet), using the genetic code to polypeptide language (protein alphabet)
5. **check validity of protein file:** protein files are built by 20 letters of the amino acid alphabet. They start with the amino acid Met.

The second project consists of using the complementary principle in genetics in order to obtain 2D spatial forms of the actors in the process of protein biosynthesis. Namely, this principle enables (among many other things) the RNA molecule to build its secondary structure. The complementary principle of the RNA molecule states that letter A is complementary to letter U (and vice versa), and the letter C is complementary to letter G (and vice versa). This principle enables the complementary substrings to be able to fold and produce a planar secondary RNA string structure. For example, the linear RNA string shown on Figure 2, can fold due to complementary forces and form a secondary structure.

Fig. 2. Linear RNA “string”

After the folding of the string given on Figure 2, we obtain a new structure, known as stem-loop, given on Figure 3.

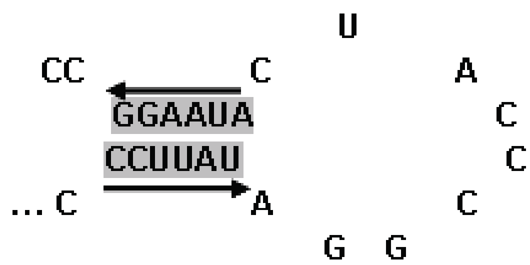

Fig. 3. Folded RNA "string" due to the complementary principle

This is used to define the second student project. Namely, the task is to find whether the RNA structure could be folded to its secondary form or not. If so, the students are supposed to locate the substrings on which the complementary principle could possibly be used. For this task, most of the students use the previously constructed module that checks the validity of the RNA string first, and then create a new module that finds substring candidates.
The third project consists of finding mRNA candidates for any given protein. Actually, it consists of creating a function that calculates the "inverse Genetic Code" - given the amino-acid sequence, calculate a possible mRNA sequence. The diversity of mRNA sequences that leads to the same protein sequence aroused discussions that life has more than one solution for the potential threats for the survival of the organism.

2.3 Robotics

Robotics for educational purposes became very popular in the previous years. Almost every curriculum of Computer science has an advanced robotics course. It is becoming so popular, that many student robotics competitions are being held worldwide. The robotics part of the Intelligent Systems course teaches students to deal with the constraints of the real world. The other subjects in their curricula are mainly covering software problems (Programming, Operating Systems, and System Software). Since the curricula on the Institute of Informatics is primarily based on software techniques, the knowledge gained when dealing with hardware is a unique opportunity for our students to gain aptitude in creating embedded software for real life applications. In this part of the IS course students encounter the problems of mass, gravitation, part assembly etc. They realize that the theoretical algorithms that are used to solve real movement need to be tuned, in order to work for physical matter.
Practical part. The approach that we are following in the practical part includes forward and inverse kinematics, velocity kinematics, vision and manipulation and biological input as robot input. We are proud to say that this approach seems to be very up to date and accepted in different universities (Wood, 2008)
For the needs of this course we used the Lynx 5 robotic arm, designed by Lynxmotion Inc (Lynx 5, 2004).
The basic idea behind these projects is to embed the software based on sensor inputs (voice and video) in movement of the robot arm Lynx 5. The projects created include chess playing

(Figure 4), video motion detection, sound forced movement, robot dancing etc. We also introduce our students to the robotics applications in medicine, especially to the Brain-Robot Interface paradigm in which recognition of an anticipatory brain potential is acknowledged and reacted to by a robot arm movement.

Fig. 4. Chess playing using the robot arm Lynx 5

2.4 Data Mining (DM)

Another part that is offered in this course, yet on a smaller scale, is Data Mining. The students are very interested to understand the main principles in knowledge discovering in databases - selection of data, preprocessing, transformation, data mining, interpretation, evaluation and, finally, discovering new knowledge. It is intriguing for the students to observe the differences between the data base queries and searching data bases for new knowledge and also to understand the difference between Machine Learning and Data Mining (Bramer, 2007). The students are accustomed to work in "laboratory environment" (clean data bases with known features of the input vectors, no duplicates in training or test data set). When they are confronted with the real world problems, which include noise and unstructured data, the students understand the whole process of gaining new knowledge, where Data Mining algorithms take the central place. We consider examples from a big mobile company (customer churn) and data from Customer Relationship Management. Also, we consider more scientific problems, like exploring the web logs, where the gained knowledge is predicting the next cash address, or personalization in offering web addresses in the person's cache.

Practical part. We offer the students two approaches concerning Data Mining. First, we use the open source package Weka, where different ML techniques are used in solving several toy problems. The students can understand the power of the non-standard ML techniques in discovering common features in large amounts of data. The students become aware of the importance of the preprocessing stage. The data for the problems that students deal with have a raw format, i.e. they are alphanumerical, with noise.

Second, we illustrate the use of associative rules as one of the studied DM techniques. We use a customer survey to decide how to optimize the configuration of a beauty store in order to achieve maximum profit.

3. Methodology for teaching the Intelligent Systems

Regardless of the topic (Machine Learning, Bioinformatics, Data Mining, Robotics), the main approach, i.e. the methodology is to confront the students with a problem (pattern recognition, prediction) and to propose an unconventional way of solving it.

We use the 'tools' of active and creative teaching, putting the student in the center of the lecture. Usually, the lecture time is divided in 3 main parts – evocation, presentation of the current topic and reflection.

The evocation part is in the beginning of the lecture, where we explore the previous knowledge about the current topic. The students' strong mathematical background is very important at this point. This way, the students become aware of the connection between the formal mathematical theory and the real-life problems and they are encouraged to think of the best way to solve them. Sometimes, in the evocation part we use the methodical technique of writing down the pre-knowledge of the current topic and we write the key points on the board, so everyone can read and eventually learn something new.

In the next part of the lecture – presentation of the current topic, we present the chosen problem to the students within these three steps:

First step – we introduce the chosen problem - specific for the certain part of the course. It is represented with a data set X and matching label set Y:

$$\begin{aligned} X &= \{\mathbf{x}_1,...,\mathbf{x}_n\} \subseteq \Re^m \\ Y &= \{y_1,...,y_n\} \end{aligned} \tag{1}$$

where $x_1, x_2, ... , x_n$ are vectors representing each data with m features, and $y_1, y_2, ... , y_n$ are the corresponding labels, depending on which class the input vector belongs to.

Second step – students think about the problem, discuss about the size of the data set, the possibilities of data representation (ML, Bioinformatics, DM, Robotics) and think about solving it in the traditional way (developing algorithm, using statistical procedures). At this point we discuss how to obtain the train/test data set. We emphasize the meaning of the preprocessing part, and gradually move toward the suitable way of presenting the data for the method that is introduced in the next steps.

Third step – we suggest methods (black-box techniques, data mining methods, dynamic programming for Bioinformatics, signal processing) that are suitable for solving the problem and explain the algorithms, i.e. the way they work. This is actually the new material that was planned to be introduced to the students for the current part of the course by presenting its theoretical background, history of its development and possible variations of the method. Sometimes, if appropriate, we explain the topic using educational movies or animations that were previously programmed.

In the reflection part of the lecture, we obtain the main conclusions about the subject at hand. We conclude the lesson by using different methodical techniques. Usually, there are some charts or tables, where we present the success of the new method compared to the other known methods used on the same benchmark data sets. It is very important for the students to understand the proper way of presenting the result of the given method (i.e. percentage of correct positive/negative, percentage of false positive/negative examples of the test data).

The theoretically explained methods are practically demonstrated on the exercises classes. The use of the methods is performed with the appropriate simulation software, using real-life data sets or known benchmark data sets for the given problem. Students get the insight of data preprocessing - preparing the training/testing data sets from the public databases. Dealing with the difficulties considering the public data repositories is very important, because data can contain noise due to the wrong interpretation of experiments, incorrect handling and storage. For example, in biological data bases, redundancy is a big problem - many entries in protein or genomic databases are members of protein and gene families that are versions of homologous genes found in different organisms.
Also, the students analyze and interpret the results of the method (post-processing).
It seems that, at this point, after practically going through the problem, most of the students develop extended interest about the subject.

4. Presentation of the outcomes

Each year (6 generations by 2009), the best students in the generation have attended the course Intelligent Systems. They develop interest in different parts covered by the course. At the end of the course, the students choose the topic they liked the most and produce a paper (an essay). This enables us to see which topic was the most intriguing for the students, and enables us to improve the course contents. It is not unusual these essays to become starting points for their diploma theses later on.
The statistics shows growing student interest for these areas of Computer Science, and many of our students decide to take Master and PhD theses in Intelligent Systems. Many of our students are now researchers in different Universities and Research Institutes abroad.
One of the most significant results was the organization of a special topic workshop named "Intelligent systems – Biological approach" (Ackovska & Madevska-Bogdanova, 2006) where scientists of different areas were presenting their work. This workshop was a chance for our students to present their own achievements in different topics studied while taking the Intelligent Systems course.
There are many requests for our students as leaders in the challenging development projects in the IT Industry. Many IT companies in Macedonia make special requests for the students that took this (specific) subject as a part of their degree. Since this course offers a wide area of topics, which are handled by different programming techniques, the students who take this course are able to cope with different and very challenging programming problems in their work environment. Many of them are already in managing positions or are leaders of Research and Development teams in the Macedonian IT Companies. Also, a great deal of our students is already working in research areas in well known foreign companies such as Philips, Nokia, Siemens and many more.

5. Evolution of the IS course and Future developments

The student interest in some of the course topics lead us to an idea of development of new elective courses in Data mining, Bioinformatics and Robotics for the students of Institute of Informatics and the course of DNA Programming for the students of Institute of Mathematics at the Faculty of Natural Sciences and Mathematics.

The courses in Data Mining and Robotics are already offered as elective courses at the Institute of Informatics.

The course in Data Mining is an extended version of the DM part of the Intelligent Systems course. The students learn the whole process of discovering new knowledge from a given data set, by learning about several DM algorithms for Classification and Prediction (Tree induction, Bayesian Classification, Linear and Multiple Regression, k-nearest neighbors) and Cluster Analysis (Partitioning, Hieratical Methods and Model-Based Clustering Methods). Also, Data Warehousing is a part of this course.

The growing student interest for the Robotics part of the Intelligent Systems course lead to the development of a new Robotics lab in our Institute of Informatics. Most of the students taking the course in Intelligent Systems are regular student researchers of the Robotics lab.

The newly constituted Robotics course teaches the Computer Science students about the basics of Robotics. This course is supported by the above mentioned Robotics Lab, supplied by manipulative, walking and mobile robots. The students taking this course learn about the assembly and control of robots and programming techniques involving material and mass control of the robots.

The course in DNA Programming (Roganovic & Ackovska, 2006) is offered to students of the Institute of Mathematics. The course is intended to help the students of mathematics in two ways. First, they learn about something completely different than their entire curriculum –foundations of genetics processes going on in a biological cell. This is very interesting for these students, due to the fact that this course is one of the few interdisciplinary courses taught at the institute of Mathematics. Second, the course is designed in a way that allows mathematics students to strengthen their knowledge in programming. Therefore, this course helps them to improve some very important features as file control and manipulation, pattern recognition in genetic sequences, as well as statistical research and error correction.

At this stage, the Institute of Informatics is creating new curricula for the Master and Doctoral Studies. This curriculum is intended to include two new modules that will cover Bioinformatics and Intelligent Systems subjects.

6. Conclusion

We have shown that introducing the concepts of Intelligent Systems course has enabled our students to broaden their perspective of the modern Information Technologies. Within this course, our students manage to understand that the real life problems can be solved by finding a good answer when the perfect answer is not possible to obtain. This is achieved by using contemporary methodology techniques of active teaching. A great deal of the course is dedicated to the practical use of the elaborated methods.

There are several newly offered elective courses in the Computer Science curriculum that have evolved form our IS course. The fact that they are very well accepted by the students is yet another confirmation of the success of our course. It also shows the importance of introducing different approaches to understanding and solving IT problems.

As an ultimate recognition of our approach to this course, we see the requests for our students as leaders in the challenging development projects in the IT Industry in our country and abroad. Many of them are already in managing positions or are leaders of Research and Development teams in well known IT Companies.

7. References

Ackovska, N. & Madevska-Bogdanova, A. (2005). Teaching Bioinformatics to Computer Science Students, *Proceedings of IEEE EUROCON 2005*, pp. 811-814, ISBN: 1-4244-0049-X, Belgrade, Serbia and Montenegro, November, 2005

Ackovska, N.& Madevska-Bogdanova, A. (2006). Workshop 'Intelligent systems – Biological Approach, http://www.ii.edu.mk/workshop/inteligentnisistemi/index.htm

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, ISBN-10: 0-387-31073-8 Springer, NY, USA;

Burges, C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery,* Vol.2 , No.2, (June 1998), (121-167), ISSN 1384-5810

Bozinovski, S.; Jovancevski, Gj. & Ackovska, N. (2002). Distributed Interactive Robotics Classroom, *Periodical in Optoelectronic Information – Power Technologies*, No. 2(4), (2002), (5-10), ISSN 1681-7893

Bozinovski, S.; Jovancevski, Gj. & Bozinovska, N. (2001). DNA as a real-time database operating system, *Proceedings of the World Multiconference on Systemics, Cybernetics, and Informatics*, Vol XV:Industrial systems, p.65-70, 2001

Bozinovski, S.; Mueller, B. & diPrimio, F. (2000). Biomimetic Autonomous Factories: Autonomous Manufacturing Systems and Systems Software, *GMD Report 115*, German National Center for Information Technology, Sankt Augustin, 2000

Bramer, M. (2007). *Principles of Data Mining*, ISBN-10: 1-84628-765-0, Springer, London, UK

Brown T.A. (2002). *Genomes*, ISBN 0-471-25046-5, Wiley&Sons, Inc., UK

Lewin B.(2000). *Genes VII*, ISBN 0-19-879276-X, Oxford University Press, NY, USA

Lodish, H.; Berk, A.; Zipursky, L.; Matsudaira, P.; Baltimore, D. & Darnell, J. (2000). *Molecular Cell Biology, 4th edition* ISBN-10: 0716740788, W. H. Freeman and Comp., NY, USA

Lynx 5, http://www.lynxmotion.com/Category.aspx?CategoryID=2, 2004

Madevska-Bogdanova, A. & Ackovska, N. (2008). Intelligent Systems for Computer Science Students, *Proceedings of International Technology, Education and Development Conference,* ISBN: 978-84-612-0192-1, Valencia, Spain, March, 2008, International Association of Technology, Education and Development, Valencia, Spain

Madevska-Bogdanova, A.; Nikolic, D. & Curfs, L. (2003). Classification of DNA sequences using modified SVM outputs, *Proceedings of 13- th International Conference of Artificial Neural Networks, ICANN /ICONIP,* pp. 114-117, ISBN:975-518-209-8, Istanbul, Turkey, Jun, 2003, Bogasici University, Istanbul

Madevska-Bogdanova, A. & Nikolic, D.(1999). Automatic Classification With Support Vector Machines In Molecular Biology, *Proceedings of Third International Conference on Cognitive and Neural Systems*, May 1999, Boston University, Boston, USA

Roganovic, D.Z. & Ackovska, N., www.ii.edu.mk/predmeti/DNAP, web site for the subject DNA Programming, 2006;

Thorsten, J., SVM light, http://svmlight.joachims.org/, 2003

Weka, open source Machine Learning Software, http://www.cs.waikato.ac.nz/ml/weka/ 2005

Wood, R.J. (2008). Robotic Manipulation Using Open-Architecture Industrial Arm: A Pedagogical Overwiev, *IEEE Robotics & Automation Magazine*, Vol. 15, No. 3, (September 2008), (17-18), ISSN:1070-9932

Technology Education and Development
Edited by Aleksandar Lazinica and Carlos Calafate

ISBN 978-953-307-007-0
Hard cover, 528 pages
Publisher InTech
Published online 01, October, 2009
Published in print edition October, 2009

The widespread deployment and use of Information Technologies (IT) has paved the way for change in many fields of our societies. The Internet, mobile computing, social networks and many other advances in human communications have become essential to promote and boost education, technology and industry. On the education side, the new challenges related with the integration of IT technologies into all aspects of learning require revising the traditional educational paradigms that have prevailed for the last centuries. Additionally, the globalization of education and student mobility requirements are favoring a fluid interchange of tools, methodologies and evaluation strategies, which promote innovation at an accelerated pace. Curricular revisions are also taking place to achieved a more specialized education that is able to responds to the society's requirements in terms of professional training. In this process, guaranteeing quality has also become a critical issue. On the industrial and technological side, the focus on ecological developments is essential to achieve a sustainable degree of prosperity, and all efforts to promote greener societies are welcome. In this book we gather knowledge and experiences of different authors on all these topics, hoping to offer the reader a wider view of the revolution taking place within and without our educational centers. In summary, we believe that this book makes an important contribution to the fields of education and technology in these times of great change, offering a mean for experts in the different areas to share valuable experiences and points of view that we hope are enriching to the reader. Enjoy the book!

Fax: +385 (51) 686 166
www.intechopen.com

Fax: +86-21-62489821