A Computational Framework For Academic Accreditation And Assessment In Higher Education (A3-HE) – Part 2 Technologies

ABOUBEKEUR HAMDI-CHERIF (1)
Computer College
Computer Science Department
Qassim University
PO Box 6688 – 51452 Buraydah
SAUDI ARABIA

(1) Permanent Address: Université Ferhat Abbas Setif (UFAS)
Faculty of Engineering
Computer Science Department
19000 Setif
ALGERIA

(1) email: shrief@qu.edu.sa, elhamdi62@gmail.com

Abstract: - In Part 1 of this paper, we described the main processes involved in academic accreditation and assessment in Higher Education (A3-HE). We have stressed the issue of heavy and tedious paperwork that characterize actual academic processes. On the other hand, both the internal self-examination undergone by institutions / programs and the external reviewing processes made by accrediting bodies are prone to errors, biases and subjective judgments because they are largely based on rules of thumb human judgments – despite the use of standards. In this second part of the paper, we propose a set of computational technologies for the enhancement of A3-HE. Emphasis is put on technologies spanning (crude) data, information, refined information including decision support, ultimately leading to the most refined and expensive piece of information, i.e., knowledge and its discovery in large and diversified databases over the Web. A human-machine interactive knowledge-based learning control system for A3-HE is our far-reaching goal. Because the A3-HE processes are too complex to be addressed by computerized systems alone, scaling up to real-life applications still require much time to reach maturation.

Key-Words: - Academic accreditation, Database management system (DBMS), Information system (IS), Decision support system (DSS), Knowledge base control system (KBCS), Data mining, Web technologies.

1. Introduction

Academic accreditation and assessment in Higher Education (A3-HE) is meant to acknowledge that an institution or program is following recognized requested quality criteria. It is a complex, lengthy and tedious set of intertwined processes involving people at different levels of responsibility within the community, costly machines, various artifacts, and more often than not, heavy clerk work and expensive resources in terms of time, money and other unquantifiable elements. In most countries of the world, academic accreditation processes are being conducted by government organizations such as a Ministry of Education / Higher Education or non-governmental non-profit agencies [4], [19]. In order to address the issue of heavy paperwork and subjectivity during the periodic self-examination, or during the reviewing process or at the final test for accreditation, we define some computational technologies that are shown to help. There are indeed many ways in which computational science can be applied in academic accreditation processes enhancements. We describe here the most important components and show how computers can be useful in this domain, although it must be also stressed that the A3-HE problem is too complex to be addressed by computerized machines alone. These latter can play a significant role towards alleviating the encountered multi-faceted issues, leading to an acceptable solution. Not pretending to delve into all the details of these highly complex and interdisciplinary issues, we suggest using a few entry points available to computational and control sciences. In Part 1 of this paper, our aim was to contribute to a unified view of the processes involved in A3-HE for accrediting bodies around the world. Now, we focus on computational technologies per se. These latter involve different levels of actions spanning all possible outcomes [20] ranging from gathering and exploitation of (crude)
data, to the generation of information, to the obtainment of refined information including decision making, and ultimately to the most refined piece of information, i.e., knowledge, its discovery and its subsequent usage for controlling the academic processes. The actual technology-oriented efforts only span minor processes within the A3-HE, e.g. [8], [10], [12] [14]. Hence this paper. In the next section we give an overview of the importance of data leading to rational decisions in A3-HE information systems. Section 3 discusses the transition from knowledge to its discovery, including Web technologies. Section 4 discusses the knowledge-based control learning system (KBLCS) as an ultimate architecture with the possible impacts the proposed technologies are thought to induce on future A3-HE. The paper ends with a conclusion summing up the main results and pointing towards some potential future developments.

2. Basic technologies

2.1 Related works

There exist automated tools for similar settings such as university rankings delivered by the Academic Ranking of World Universities ARWU [http://www.arwu.org], or Times Higher Education University Rankings (THE [http://www.timeshighereducation.co.uk/], or Webometrics Ranking of World Universities (WRWU, [http://www.webometrics.info/], to cite but a few. There are also automated tools certification and assessment for some professions [5]. In Industry, surveys confirm that certification and accreditation (C&A) approaches in private and government sectors differ widely. However, they now share a general trend towards automated processes replacing costly manual approaches. In the US Department of Defense (DoD), the Defense Information Technology Systems Certification and Accreditation Process (DITSCAP) and the newly-established Department of Defense Information Assurance Certification and Accreditation Process (DIACAP) guides C&A and specialized information systems. A study has been undertaken that provides an analysis of inflexible and flexible design approach costs of both automated and the currently employed C&A manual processes prescribed by policy. The study deals with the Global Positioning System Joint Program Office (GPS JPO) [7]. These efforts, although too specific can provide useful guides for the development of corresponding future A3-HE computational frameworks. The main issue to date is that there is no such counterpart for A3-HE.

2.2 From data upwards

We stress the stratification of A3-HE processes into levels. This stratification motivates for the introduction of incrementally-sophisticated technologies. The proposed solution to the issues addressed in A3-HE settings is to be considered under six hierarchical and complementary levels, namely data base management system (DBMS), information system (IS), decision support system (DSS), knowledge-based system (KBS), data mining (DM) and knowledge base control system (KBCS). The main aim is the integration, within A3-HE, of learning methods, encompassed by the knowledge base learning control system (KBLCS) framework.

2.3 Overview of ad hoc technologies

The knowledge-based learning control system (KBLCS) layers are depicted in Figure 1. The proposed hierarchical study spans gathering and management of data, generating information, producing knowledge, discovering knowledge, then controlling A3-HE processes using human-machine interaction, proactively. In order to address the A3-HE processes from a computational perspective, we adopt a bottom-up approach, i.e. starting from the simplest to the most complex technologies. Because we are interested in architectures and not in specific implementation, some of the proposed technologies, considered here, are still in their early infancy and did not reach sufficient maturation for application in simple contexts, let alone in A3-HE settings. The driving technology in the present approach is knowledge base control system (KBCS), considered as the core of knowledge base learning control system (KBLCS). There exist several feedback loops. The main ones are the outer feedback based on the interaction with Web technologies, including pervasiveness, based on mobile-like technologies and cloud computing. Machine learning processes (ML) and common sense acquisition constitute the middle and inner feedback loops, respectively. The results given by these technologies integration are meant to be used by all parties, i.e. by institutions or programs seeking A3-HE status, by decision makers at different levels of accountability including government policy and legislature makers.
2.4 Database management system (DBMS)

Any A3-HE project or implementation has to start with the most basic element *i.e.* crude data. In any A3-HE context, data is very heterogeneous. For instance, at the most basic level, crude data might consist of number of students per year in a given program who have completed/dropped courses, with their grades, courses with major failure rates, faculty completing course schedules as planned, the number of highly qualified faculty wishing to join/leave the course/institution, the impact of the program on employers, to cite but a few scattered pieces of data. In most cases, data has to address at least curricula contents, students, faculty, staff, and employers. As the volume of data grows, it becomes necessary to use a database management system (DBMS), *i.e.*, a coherent computational environment that controls the creation, use, and the maintenance of integrated collection of different data records and files known as databases. A DBMS allows end-users and other software to store and retrieve data in a friendly and structured way. Indeed, as data becomes more and more diversified, the techniques used to design and implement DBMSs have become more sophisticated to cater for large-scale heterogeneous data and noise [6]. Instead of having to write computer programs to extract information, most systems offer the possibility of using special query language or fourth-generation programming language (4GLs) along with other application development features. Traditional DBMSs whether they are relational, object-based, network, hierarchical, or multidimensional, are considered as intelligence-free programs. By intelligence-free program, we mean a program that does not infer new conclusions not explicitly stored in the computational environment.

2.5 IS and DSS

A series of technical advances in recent years has increased the amount of data that decision makers can record about different characteristics of any academic enterprise and its by-products such as students, research contributions, and community involvement or impact on it. It has to be noticed that any academic institution is literally flooded in a deluge of data. This data is, of course, vital for advancing our appraisal of the enterprise. Despite the fact that DBMSs are unavoidable in any A3-HE setting, they are far from being sufficient. Indeed, on top of DBMS technology, we need a more structured information as a form of processed data. This naturally leads to the use of information system (IS) [15]. On the other hand, there is no proper decision without information. Because information is becoming more and more sophisticated, proper decisions are becoming more and more difficult to take. This issue is handled through decision support system (DSS) For instance, if we consider an institution or a program as an enterprise evolving in a competitive marketplace, then it becomes obvious that it has to consider constantly-changing external parameters to stay alive and competitive. For that, information system (IS) and decision support system (DSS) technologies are necessary.

2.5.1 Information system (IS)

Let us note at the outset that information system (IS) is not to be confused its own discipline - information systems. In a very broad sense, the term information system (IS) refers to the interaction between people, algorithmic processes, data and technology. The term refers not only to the information and communication technology (ICT) within an organization, but also to the way in which people interact with this technology in support of the organization’s business processes [15]. As a result, IS involves both the infrastructure and the infrastructure that supports it [9].

2.5.2 Decision support system (DSS)

When exactly to take a given decision within the institution or program? What precise items are to be changed within the institution/program? What curriculum to adopt to face fierce competition and market trends? These issues are addressed within a decision support system (DSS) framework. A typical DSS consists of three principal components: a data warehouse server, online analytical processing (OLAP) and data mining tools, and back-end tools for populating the data warehouse. The Data warehouses contain data consolidated from several operational databases and tend to be orders of
magnitude much larger than operational databases. Typically, the data warehouse is maintained separately from the organization’s operational databases because analytical applications’ functional and performance requirements are quite different from those of operational databases. Data warehouses exist principally for decision support applications and provide the historical, summarized, and consolidated data more appropriate. Due to the fact that data warehouse construction is a complex process that usually takes many years, perhaps decades, we can instead build data marts, which contain information for specific departmental subsets [3]. For example, an alumni follow-up data mart may include only type of degree, time of employment after graduation, and the employer. Several departmental data marts can coexist with the main data warehouse and provide a partial view of the warehouse contents.

2.5.3 Example of DSS/A3-HE interaction
As an example of DSS used for accreditation, we can cite the AEFISTM (Academic Evaluation, Feedback, and Intervention System http://www.goaefis.com/about-aefis/faqs/). AEFIS is a web-based assessment management solution and Instructional Decision Support System (IDSS) that supports personalized learning by delivering real-time information to administrators, faculty, students, and alumni and industry to enhance curriculum development and help in the accreditation processes. The core of AEFISTM 3.0 is the so-called Instructional Decision Support SystemTM (IDSSTM). It is a web-based system that gathers information to help instructors make educational decisions based on student data to improve instructional practices. However, AEFISTM has no machine learning layer.

3 From knowledge to its discovery
This section gives the basic outlines for addressing two fundamental issues. First, it investigates the contribution of data mining technologies to A3-HE for the discovery of hidden patterns in data. For that, we describe the computational contributions to A3-HE and show how it can help in developing increasingly-complex tools for addressing knowledge discovery and issues. Second, we consider control issues within the framework of knowledge base system (KBS) and machine learning frameworks, ending with knowledge base learning control system (KBCLS) approach.

3.1 Knowledge-based system (KBS)
Knowledge can be considered as the one of the most important asset in our modern society and particularly in academia. It is omnipresent in all endeavors of computing practice: from the rise of knowledge management to the Semantic Web and from the new blog culture to the knowledge economy. This penetration has made proper knowledge base systems (KBSs) and their engineering one of the most requested features. A KBS is a specialized software that uses artificial intelligence (AI) methods such as expert system in problem solving processes. It is composed of a rule base, a fact base and an inference engine. The most precious part of any KBS is its of expert knowledge embodying human knowledge, while a typical inference engine can be obtained freely from the Internet, e.g., the “C Language Integrated Production System” (CLIPS http://clipsrules.sourceforge.net/). A KBS has also couplings and is designed to facilitate knowledge retrieval in response to specific queries, or to transfer expertise from one domain of knowledge to another.

3.2 Data mining
3.2.1 Defining data mining field
Data without the pertinent knowledge that conveys it might be useless. Knowledge can be seen as the patterns or characteristics of the data. Raw data is sometimes meaningless because what we want is the knowledge hidden in the data and not the data as such. That is why a new technology has emerged in the mid 1990’s to deal with the discovery of knowledge from data. It is called knowledge discovery in databases (KDD) or simply data mining (DM). Uncovering hidden information is the fundamental goal of data mining. A distinctive aspect of actual data mining systems is their extensive use of the Web and the manipulation of diversified data. In A3-HE, data might be constituted of student or faculty retention rate over a given period of time within a program or institution. The reasons behind these retentions rates constitute the knowledge extracted from DBMSs using specialized methods such as classification, clustering or association. Now some freely-available platforms implementing most data mining algorithms are available on the Web, e.g. Tanagra [http://eric.univ-lyon2.fr/~ricco/tanagra/], and Weka [www.cs.waikato.ac.nz/ml/weka/].
3.2.2 Data mining process and tasks

(i) Process

Data mining process is based on the following steps: data collection, data preprocessing, data mining proper, information interpretation, and visualization.

(ii) Tasks

The data mining tasks are categorized as follows.

- **Classification** decides the class/group for each data sample. For example, students withdrawing from a given program can be classified based on their pre-secondary schools.

- **Clustering** is to regroup similar data into a finite set of separate clusters/categories. This task is also referred to as segmentation. In machine learning, it requires unsupervised learning i.e. that the clusters are not known in advance.

- **Association**, **link analysis** or **affinity analysis** is to tell whether a set of data depends on the rest. An association rule can be written as $A \rightarrow B$ where both $A$ and $B$ are a data set.

- **Summarization** or **characterization** is a simple description of a large data set. It is desirable that representative information of a data set can be obtained, so that we can easily have a general view of the data set.

- **Text mining** is used when the data to be mined are text instead of numerical data. It originated from information retrieval (IR) of the library science. Keywords are used to find related documents from a document database. For more advanced applications of text mining, classification, clustering, and association techniques are utilized. Text mining can facilitate the re-examination of the literature to link the facts that were already known [13]. In A3-HE, text mining can be used in the analysis of reports written in the self-examination by institutions or by accrediting bodies.

3.3 A3-HE issues and data mining

In A3-HE, knowledge discovery is of utmost importance. Computer Science program coordinators and curriculum committees are confronted to a dilemma. On the one hand, the organization and content of a scientific discipline should not depend on the whims of employers alone. On the other hand, the new technological environment is under the pressure of globalization and contains forces beyond the control of governments, let alone curriculum committees, and marketplace effects are forcing academic environments in terms of enrollment and faculty positions [11], [16] . Data mining is a framework that helps in alleviating this dilemma by making a rapprochement between academic and professional aspects of the issue.

3.4 Web technologies

3.4.1 From Web 1.0 to Web 2.0 technologies

All the previous issues dictate the urge for further integration of multiple technologies such as Web 1.0 / Web 2.0 technologies. There was a natural transition to Web 2.0 after the so-called dot-com crash in Fall 2001. The main Web 2.0 technologies include wikis, blogs, RSS filters, folksonomies, mashups, podcasts, crowdsourcing, social networks, and virtual worlds [2]. A summarized comparison between Web 1.0 and Web 2.0 technologies relevant to A3-HE is given in the Table 2 [http://oreilly.com/web2/archive/what-is-web-20.html], [17].

For example, use Web 2.0 technology in A3-HE to identify degree and accreditation mills through social networks and blogs, to cite but a few.

<table>
<thead>
<tr>
<th>Web 1.0</th>
<th>Web 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Britannica Online</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>personal websites</td>
<td>blogging</td>
</tr>
<tr>
<td>domain name speculation</td>
<td>search engine optimization</td>
</tr>
<tr>
<td>page views</td>
<td>cost per click</td>
</tr>
<tr>
<td>screen scraping</td>
<td>Web services</td>
</tr>
<tr>
<td>publishing</td>
<td>participation</td>
</tr>
<tr>
<td>content management systems</td>
<td>wikis</td>
</tr>
<tr>
<td>directories (taxonomy)</td>
<td>tagging (“folksonomy”)</td>
</tr>
<tr>
<td>stickiness</td>
<td>syndication</td>
</tr>
</tbody>
</table>

Table 2 - Web 1.0 vs. Web 2.0

3.4.2 Example of DBMS/Web 1.0 interaction

At a higher level of the A3-HE processes, data might consist of the codification of good practices as proposed by institutions for discussion / amendment / adoption and might address issues related to market trends, constraints imposed by legislatures, globalization and others.

As an interaction between DBMS and Web 1.0 technologies, we can cite the so-called “good practice database” that was developed and is now managed by the Australian Universities Quality Agency (AUQA). The Database is intended to gather information about good practices in the Australasian Higher Education sector publicly available. It is intended that such information will serve as a general guide to the sector to assist it in its efforts to improve. The Database is not intended to present standards for the sector nor will AUQA audit an institution/agency against entries in the Database.

A new batch of Good Practices was launched in September 2010, including entries from seven Higher Education institutions. Submissions covered a range of topics and now with over 200 entries, the database provides an important tool to benchmark institutional practices. Invitations for batch 14 have
been distributed and accepted with 18 good practices currently in progress. This batch is due to be launched in February 2011 and will include a number of contributions which will assist the sector to compare practices among its institutions. (AUQA http://www.auqa.edu.au/gp/terms.php).

3.4.3 Example of DSS/Web 2.0 interaction

Consider the following decision-based issue. Is a given program/course worthwhile teaching? This issue is reduced to the following: What Web 2.0 technologies say? An important part of the answer can be supplied by wikis, blogs, crowdsourcing, social networks, involving relevant community members such as faculty, employers, students, professionals, parents, and the general public. Web technologies can help in shaping the final decision by encouraging, accepting, amending, or discouraging such a program/course.

3.4.4 The advent of cloud computing

Another major technology that is susceptible to address major A3-HE issues is cloud computing. This novel technology provides a way to develop applications in a virtual environment where computing capacity, bandwidth, storage, security, and reliability aren't issues because users don't need to install the software on their own system. In a virtual computing environment, users can develop, deploy, and manage applications, paying only for the time and capacity used while scaling up or down to accommodate changing needs or external requirements. Cloud computing has the following characteristics:

(i) Distributed applications

The storing of, and accessing to applications and computer data often through a Web browser rather than through specific software installed on every personal computer or server.

(ii) Pervasiveness

Internet-based computing whereby information, IT resources, and software applications are provided to computers and mobile devices on-demand.

(iii) Service-oriented application (SOA)

Using the Internet to access Web-based applications, Web services, and IT infrastructure as a service. The applications (e.g. the email service Gmail [http://www.gmail.com]) live on a remote “cloud server” that anyone can access through the browser like any other website. Rather than every user having their own copy to install maintain and upgrade, a single cloud application at a single location is shared by millions of users, or more, on a daily basis. (http://www.cloudcomputingdefined.com/learn/).

Although cloud computing is a relatively new technology, some solutions have appeared recently. For instance, IBM™ has partnered with Amazon™ Web Services to give users access to IBM software products in the Amazon™ Elastic Compute Cloud (EC2) virtual environment. This is product-level code with all features and options enabled. http://www.ibm.com/developerworks/cloud/ec2.html

3.4.5 Cloud computing and A3-HE

A3-HE large databases surely contain interesting patterns waiting to be discovered. Data mining, as expanded above, is a suitable technology helping in this discovery. On the other hand, cloud computing, through interaction with human experts can lead to novel theories about academic processes and their modeling/control entailing the possibly of designing ad hoc tools for addressing some of the A3-HE pending issues.

4. A control approach to A3-HE

Both human-made and pure human systems exhibit collective behaviors amongst individuals in which some controllers seek a certain form of equilibrium. Based on this idea, we propose a framework whereby the A3-HE is considered as a control system.

4.1 Control theory viewpoint

In control theory, a controller is a device which monitors and affects the operational conditions of a given dynamical system. The operational conditions are the output variables of the system which can be affected by automatically adjusting certain control variables. In our situation, the A3-HE processes are equipped with a controller for measuring the output variables, i.e. A3-HE outcomes which are feedback for corrective action. In this example, the assessment is the controller and directs the activities of the A3-HE processes. The actions taken for improving the outcomes are the desired outcomes, or setpoint. In agent-based terminology, the A3-HE is viewed as a multiagent entity producing actions (actual outcomes) under the influence of inputs (objectives, specifications) within a certain environment (community) and in which a set of human expertise is coded so that the difference between the actual outcomes and desired ones are kept to a minimum.
4.2 KBLCS
As far as knowledge is concerned, the hierarchical control processes consider two layers. The first is knowledge base control system (KBCS) and the second is knowledge base learning control system (KBLCS).

4.2.1 KBCS for A3-HE
Intelligence-based or AI-based programs are the basic characteristics of KBCS. Indeed, for complexity reasons, ‘brute force’ use of procedural heuristics-free algorithms is useless. Clearly, because of the nature of any A3-HE context, a more ‘intelligent’, i.e. heuristics-based approach is required to solve problems since human expertise codification is necessary. KBCS principles and tools enhance the A3-HE existing solutions because it incorporates human expertise in the form of code. As to learning control, it is traditionally linked with machine learning technologies such as fuzzy systems, neural networks, genetic algorithms, among others [21].

4.2.2 Machine learning
Machine learning is an adaptive process that makes computers improve from experience, by example, by analogy, or otherwise. Learning capabilities are therefore essential for automatically improving the performance of a computational system over time on the basis of previous history. A basic learning model typically consists of the following four components:
- learning element, responsible for improving its performance,
- performance element, or decision support system (DSS) responsible for the choice of actions to be taken, whether these actions are decisions or controls,
- critical element, which tells the learning element whether the criteria are met within some critical boundaries,
- problem generator, responsible for suggesting actions that could lead to new or informative experiences [18].

The importance of learning will grow as the external environment continues to generate and integrate large quantities of diversified data.

4.2.3 Supervised vs. unsupervised learning
As far as relevant machine learning is concerned, there are basically two categories of learning schemes, namely supervised learning and unsupervised learning. Supervised learning learns the data with a known answer at hand. From the control standpoint, this is a typical feedback control system. The parameters are modified according to the difference of the actual output and the desired known output, or expected answer. Classification falls into this category. On the other hand, unsupervised learning learns without any knowledge of the outcome. Clustering belongs to this category. It finds data with similar attributes and aggregates them in the same cluster [1].

The main familiar machine learning methods such as decision trees learning (DTL), and support vector machines (SVM) have proved very useful in addressing both classification and clustering problems. But machine learning techniques usually handles relatively small data sets because the learning procedure is normally very time-consuming. To apply the techniques to data mining tasks, the problem with handling large data sets must be overcome [13].

4.3 KBLCS impacts on A3-HE
We believe that the integration of previously-described technologies will advance our knowledge of A3-HE processes based on the most powerful technological tools available to computer and control systems scientists, entailing a better modeling and control of the A3-HE processes. The impacts on many fields of research are expected to be important, not only on academic processes as such but also on the community as a whole. We expect impacts on the following fields of research and technology.

i. DBMS: More structured organization of data for efficient response to queries. For instance, to develop ways to index or otherwise preprocess the data to make searches more efficient.

ii. IS/DSS: To provide intelligent interfaces that will assist the user in framing and executing queries.

iii. KBS/KBCS/KBLCS: To further formalize A3-HE within knowledge base systems (KBSs) and engineering frameworks with machine learning capabilities.

iv. Web technologies: To further integrate A3-HE within actual Web/cloud computing technologies and solutions.

5 Conclusion
Academic accreditation is above all a recognition by the community that a given academic institution meets some expected quality standards and generates recognized quality products such as students, research production, and services. We have shown how to proactively increase the collaboration between human decisions and computerized A3-HE processes. On top of the multitude of methods and
tools, it remains highly expected that the technologies reported here will uncover even more useful hidden structures in A3-HE processes. In addition to actual query methods now available, however intelligent these might be, future public A3-HE databases have to include an array of “what-if” simulation scenarios capable of producing highly anthropomorphic behavior through the knowledge base learning control system framework. Further integration of these technologies into A3-HE will remain indeed a challenging task for years to come.

References:


Websites accessed as of January 2011.

CLIPS: http://clipsrules.sourceforge.net/
Cloud computing: (http://www.cloudcomputingdefined.com/learn/)
THE – http://www.timeshighereducation.co.uk
WRWU, http://www.webometrics.info/
Web 2.0 http://oreilly.com/web2/archive/what-is-web-20.html