Designing a Modern IT Curriculum: Including Information Analytics as a Core Knowledge Area

Magnus Westerlund
Department of Business Management and Analytics
Arcada University of Applied Sciences
Jan-Magnus Jansson in aukio 1
00560 Helsinki, Finland
magnus.westerlund@arcada.fi

Göran Pulkkis
Department of Business Management and Analytics
Arcada University of Applied Sciences
Jan-Magnus Jansson in aukio 1
00560 Helsinki, Finland
goran.pulkkis@arcada.fi

Abstract
Much has happened in the field of Information Technology since 2008 when ACM published its curriculum recommendation for a four year Undergraduate Degree Program in Information Technology. We show an alternative path reflecting what we consider presently requested by the industry and students alike. In this paper we look at the topics from a holistic point of view, not just as traditional machine learning Computer Science courses. We make an argument for widening the scope from machine learning theory, towards analytical service development. We give our proposal of a refined IT curriculum that can be used by other institutions for refining their curriculums.

Keywords:
IT2008, Model Curriculum, Information Technology Education, Information Analytics

1. Introduction
Technology development in the IT sector has during the previous decade, to a large extent, focused on software service development. This has allowed the industry to open up previously closed systems toward information sharing modules. However, during recent years we have seen a new trend emerge, which seems to become the main driver for the IT field in the foreseeable future. In their search for improving customer offerings and increasing productivity, companies and other organizations today turn towards analysing data and information by using advanced machine learning models. The development and usage of these advanced model types have previously mostly been part of master and doctoral Computer Science studies, but we argue that this is about to change. Currently many cloud service providers are developing mainstream offerings of machine learning services that can rapidly be implemented in any software service offering. In consequence, this will require a different skill set compared to what most IT engineers have been taught up till today.

The field of Information Technology undergoes rapid change and requires teaching organizations to continually redevelop their curricula to reflect the changes in the field. Current jobs of professionals with an undergraduate degree in Information Technology (IT, we refer henceforth to the academic discipline and not to the field) are quite often more closely related to the business side of an organization than the jobs of Computer Science (CS) professionals. CS professionals use their scientific competence to solve technical problems and to design software, devices, and systems. An IT professional may, in addition to working on similar tasks, need to understand and communicate sometimes complex dependencies or abstractions between CS professionals’ scope and business oriented clients.

With the advent of Information Analytics and software services in general (Software as a Service, SaaS), we see some new trends in the role of IT professionals in the future technology landscape. In the future an IT engineer must also be able to communicate technical implementations, related to information analysis, to business stakeholders. The volume of data and information is growing rapidly in the operational environment of business organizations and other organizations. IT engineers must in the future also have Big Data processing competences such as Information Analytics, which requires profound machine learning skills. We define Information Analytics as a broader knowledge area than Business Analytics that primarily considers business information. Information Analytics deals with all types of information and focuses on the creation of software solutions and services that process this information.

The remainder of this paper is structured as follows. Section 2 contains a literature review of related curriculum design research. Section 3 defines the knowledge areas and learning methods of a proposed new IT curriculum. Section 4 describes the competence requirements of the modules in the proposed new IT curriculum. Concluding remarks are expressed in section 5. Details of the proposed new IT curriculum are shown in an Appendix.

2. Related Curriculum Design Research
Current curriculum design research is usually related to competencies, which the students should achieve. The following five characteristics

1. inclusive and integrative,
2. combinatorial,
3. developmental,
4. contextual, and
5. evolutionary
are used in (Tardif 2006) to describe the features of competency.

In Chang (2014) a method called Q-technique is proposed to obtain the competence requirements of IT enterprises to serve as a basis for developing an IT curriculum:

“the purpose of the Q-technique is to consult leading experts within the IT industry, to obtain statements and priorities, to form the universal requirements of IT professional competency.”

As a conclusion is stated that

“IT competencies are structured along the dimensions of information ability, fault tolerance, execution ability, problem solving, learning ability, and innovation ability.”

A process workbook for implementing competence based education has been prepared for the Clinical and Translational Science Institute of the University of Pittsburgh (Dilmore, Moore, and Bjork 2011). Design of a curriculum in a chosen discipline consists of 13 generic process steps and 3 generic implementation steps.

Design of a competency based curriculum content framework of mechanical technology education is presented in (Sudsomboon 2007). The framework is a set of requirements for knowledge and understanding, for skills, and for attitudes.

ACM currently gives curriculum recommendations for Computer Science (CS), Computer Engineering (CE), Information Systems (IS), and Software Engineering (SE), in addition to the recommendation for Information Technology (IT) (ACM 2008). Examples on using the ACM IT2008 recommendation (Lunt, et al. 2008) for IT curriculum design are found in (Koohang, Riley, Smith, and Floyd 2010, Adegbehingbe and Obono 2012).

Only minor changes were made in the ACM Computer Science curriculum recommendation for 2008 in comparison with the recommendation for 2001, while the changes are significant in the recent recommendation for 2013 (Computer 2013) in comparison with the recommendation for 2008. Required study time in the hardware related knowledge areas ‘Architecture and Organization’ and ‘Networking and Communication’ has been cut more than 50% and two new knowledge areas, ‘Information Assurance and Security’ and ‘Parallel and Distributed Computing’ have been added. In ‘Architecture and Organization’ there is a stronger emphasis on multi-core parallelism and virtual machine support, and in ‘Networking and Communication’ there is increased attention to wireless networking. The required study time in ‘Information Assurance and Security’ is more than 20% of the total required study time when the distribution of this knowledge area in other knowledge areas is taken in consideration.

The significant changes in the Computer Science curriculum recommendation for 2013 (Computer 2013), in comparison with the recommendation for 2008, strongly motivate large changes in the ACM IT2008 Model Curricula recommendation in (Lunt, et al. 2008). ACM Education Board has actually in 2012 established an exploratory Review Task Group for Information Technology (RTGIT) to review the IT2008 curriculum recommendations (Paterson et al. 2013). In Zilora et al. (2013) a new curriculum proposal for teaching IT and also the addition of analytics as an overarching theme for the curriculum is presented.

3. Knowledge Areas and Learning Methods of a new IT Curriculum

When developing our proposed curriculum we primarily make our recommendations for change based on the experiences in research within generative information infrastructures (Henfridsson and Bygstad 2013), the three branches of analytics (descriptive, predictive, and prescriptive analytics, outlined for business analytics in Delen and Demirkan (2013), but also applies generally for information analytics), natural computing (Shiffman 2012), software engineering, and modern pedagogy of integrating natural sciences and programming through an Active Learning methodology.

3.1 Knowledge Areas

We started our curriculum development by considering whether we should merely add one or two more pillars to the existing five pillars of IT, presented in the IT curriculum recommendation in Lunt, et al. (2008). The IT field has however changed tremendously over the past six years. By just adding more pillars we feel that the curriculum would become too general and would therefore be insufficient for IT students and for industry requirements on future IT experts. We concluded that a clear focus is required in order to ensure the necessary depth in the education of IT engineering experts. We propose that an IT professional should acquire competences in four different knowledge areas, which also reinforce each other, see Figure 1. However, we should point out that the current broader competence, based on the ACM IT2008 Model Curricula (Lunt, et al. 2008), is still relevant today. With the introduction of cloud computing services like PaaS (Platform as a Service), SaaS (Software as a Service) and IaaS (Infrastructure as a Service) over the past years, we conclude that in the future there will be fewer jobs requiring hardware knowledge and to some extent hardware related networking skills. Most new IT jobs will require software skills related to IT services and scalability issues. A good example of this development is the Amazon AWS Elastic Beanstalk (AWS 2014) solution for implementing Java applications to a publicly available application server without any in-depth hardware knowledge. We believe that it is therefore essential to look towards the future and to do our best as educators to anticipate coming industry needs.

As the IT curriculum proposal in Zilora et al. (2013), our IT curriculum proposal also has a focus on analytics. It consists of the following 8 modules:

- General Studies
- 4 Core technical modules
  - a basic study module on Web and Visualization
  - professional study modules in Analytical Methods and Data Science, Service Oriented
The main focus for all natural science topics throughout the degree is to provide students with a fundamental mathematical understanding of machine learning and data science relevant concepts. We realize this through the perspective of natural computing significant course units. This, we think, assists the student in the learning processes by providing a reference model that the student can relate to, by seeing how nature functions.

During the first year, students will study natural science course units for a total of 15 ECTS. The studies focus on applied mathematics, statistics, physics, and introduce them to mathematical programming from the onset. To give one example, physics will be taught in the form of game programming in order to help the student to visualize essential concepts. The intention is to positively reinforce the students’ learning process, to focus the students on their own experience and development. We, as in most Western European societies, have noticed a decline in natural science ability and interest among the recent generation of students. We hypothesize that students will become more motivated if the focus in natural sciences is not only on abstract topics, but also involves creative and responsive elements.

Throughout the second year students learn to understand the concept of particle systems, as a collection of independent but interactive objects (Reeves 1983). The students should be able to implement a system of particles interacting based on forces, motion, waves, and oscillations, in order to understand the notion of variations over time. Concepts such as amplitude, frequency, period, degrees and radians and their transformations become familiar, e.g. in programming a pendulum example (Shiffman 2012). An important part is matrix calculations, including scalar and matrix operations, transverse matrix, inverse matrix, determinant and solution of matrix equations.

4. Competence Requirements for an IT Engineering Curriculum, Focusing on both Information and Technology

In this section we start by describing the general competences we have found to be essential for the future IT engineer. This is followed by an examination of the core technical competences required to handle the diverse responsibilities. During this examination we do not limit ourselves to Information Analytics, but rather present a holistic view of the core technical competences that were identified. In essence we define the future competence need to be based on equal parts of Information and Technology studies. The level of competence depth gained for the degree, i.e. learning outcomes, should follow the European Qualifications Framework (European 2014) level 6. Level 6 competences are defined in the context of responsibility and autonomy as “Manage complex technical or professional activities or projects, taking responsibility for decision making in unpredictable work or study contexts; take responsibility for managing professional development of individuals and groups.”
4.1 General Competences
As the Saas model has become an important driver for digital entrepreneurship and business growth, we believe students need an in-depth understanding of creating such services. Service innovation is mostly a customer driven process so customer involvement is important (Hanseth et al. 2012, Alam and Perry 2002). Understanding and being able to anticipate the latent needs of customers is a complex task, but research shows that customer involvement is often crucial and leads to the development of more innovative services, regarding both originality and user value (Matthing et al. 2004). Therefore it stands to reason that communicational, social, and business skills among IT engineers are of great importance. We consider that this is becoming an even more prominent feature of successful IT professionals in the future. In Lunt, et al. (2008) it is stated that IT “focuses on preparing graduates who are concerned with issues related to advocating for users and meeting their needs within an organizational and societal context through the selection, creation, application, integration and administration of computing technologies”. In our IT curriculum proposal the studies will therefore contain a significant amount of general competences in topics such as communication and social interaction.

In Delen and Demirkan (2013) the need for business analytics is defined as “At a time when firms in many industries offer similar products and use comparable technologies, business processes are among the last remaining points of differentiation.” The field of analytics allows the companies to extract “every last drop of value from those processes”. This requires at least a basic understanding of how a company functions and how different business processes can be measured. We therefore defined an elective module as business processes, were students can get an introduction into functionality of companies. We introduce three different core processes: marketing, logistics and financial management.

We also offer an alternative to this module, an entrepreneurship focused module for those students that prefer founding their own companies.

4.2 Overview of Core Technical Competences
In designing the curriculum the outcome goals for the degree were defined as that graduates have competences to analyse information and are able to develop software services for the digital world, here without focusing on any specific context area. The student should learn to plan and construct software for web, mobile, and cloud services or applications. The student should also have the ability to visualize, analyse, and handle data that exist in various forms. The final technical goal that was formulated was that the student should be able to motivate the use of different types of machine learning models in order to get answers from various hypotheses or to questions based on processed data.

We acknowledge that Big Data has become an important impetus for many technology oriented and customer driven companies. However, we find the analytical understanding from an academic perspective to be the fundamental driver for implementing new services.
methods in scalable architectures. With the introduction of Big Data tools such as Hadoop we have gone from mostly focusing on efficient data structures to designing efficient algorithms that process seemingly unstructured data. Hadoop (Welcome to Apache Hadoop 2014, Borthakur 2007) can be described as a distributed database that allows users to process information directly on the node were the data resides. Therefore understanding the concept of parallelization for solving problems becomes essential.

Before a dataset can be processed in a machine learning model it often requires extensive pre-processing, also called feature engineering. There are several stages that are part of pre-processing, e.g. checking data for validity, coding, dealing with missing values, normalization, and feature extraction, to name a few. The pre-processing stage is usually considered to be the most time consuming and important stage in terms of improving the end result. In addition to understanding the earlier mentioned stages it also requires object oriented programming skills in order to be able to automate data pre-processing. Once a machine learning model has given an output, this data needs to be post-processed to create decisions and often some type of visualization is needed for a human to understand the output. (Baesens 2014)

Taking this into account we have defined the targeted competences for the module to be that the student can manage, organize and visualize data. The student should also be able to justify how the data should be stored to comply with technical, legal and contractual provisions, but also be able to evaluate security risks in data management and apply data security in computer networks. Regarding programming competences the student should be able to plan and produce secure applications based on object oriented programming. Students learn to develop essential sequentially coded algorithms and also to implement these algorithms with parallelized code. Examples of such algorithms are various sort/process tasks, which in later modules can be used to explain more complex programming methods, e.g. the MapReduce programming model (Dean and Ghemawat 2008).

4.5 Service-Oriented Architectures and System Design Module

As the software industry has matured over the past two decades it has meant that we currently emphasize architectural design more than ever. The focal point for this development has been the Service Oriented Approach which represents a baseline for a distributed architecture with no direct reference to implementation (Erl 2004). The distributed architecture defines well-formed access points through Web services, which when made public, open up the information infrastructure to become a shared, evolving, and open experience (Hanseth 2002). Henfridsson and Bygstad (2013) identified three generative mechanisms at the core of creating successful information infrastructures: innovation, adoption, and scaling. These were considered self-reinforcing processes that spawn new recombinations of resources. As user adoption increases, more resources are invested and therefore the usefulness of the infrastructure increases. True service scaling attracts new partners by offering incentives for collaboration.

We consider it important that students understand that creating successful software requires a much broader understanding than only a programming understanding. Hence, we will devote a large portion of the third year towards raising awareness and understanding of how software architectures can be made scalable by utilizing cloud infrastructures and software defined networks (Sommerville 2013). As a basis for an innovative infrastructure we will focus the attention on how students obtain a critical understanding of descriptive data mining or text analytics. From a technical point of view the student must be able to defend the chosen architecture by referring to established software patterns.

The objective for the third year is to give an introduction to machine learning models. The core focus is on autonomous agents, evolutionary algorithms, and statistical pattern learning (text analytics). Students learn to design ranking algorithms that allow them to implement objective functions for various optimization problems. Earlier studies have focused on the individual particle, but during the third year “herd behaviour” is introduced as to appreciate how the agents’ own decisions influence the group and vice versa.

As is mentioned in section 4.1, the recommended general competences include topics such as marketing and digital marketing to enhance the students’ understanding of consumers and service adoption.

4.6 Machine Learning and Decision Support Systems Development Module

In Davenport (2013) it was claimed that we are currently embarking on the third evolution of analytical services. Analytics 1.0 was the era of Business Intelligence, 2.0 the era of Big Data, and 3.0 is the era of data-enriched offerings. This new era requires new types of technologies, but also uses many of the open source tools, e.g. Hadoop, or cloud computing services developed for the previous eras (Davenport 2014). Therefore we claim that the new IT engineering challenge will be to combine various tools and services with appropriate models and data sources, to deliver new insights to the end user.

Thus, the fourth and final technical module focuses on Service Oriented Decision Support System (SODSS) development. These service types are often offered as distributed collaboration components, produced by many partners, and consumed by end users for decision making. Examining the SODSS environment as a process three major service classifications emerge: data, information, and analytics. Data-as-a-Service (DaaS) allows any business process to access data wherever it resides. The technical implementation is often performed through Master Data Management (MDM) and/or Customer Data Integration (CDI). Information-as-a-Service (IaaS) typically refers to a refinement of data and to making information available quickly to people, preferably in real-time. This opens up technical challenges such as real-time data formatting, in-memory computations, and parallel transaction and event processing. Analytics-as-a-Service (AaaS) tends to focus on insights drawn from machine learning models. These models can be of a descriptive nature, but are often focusing on predictive and
prescriptive elements. AaaS consumes information services in order to deliver different types of Enterprise Analytics or other end user relevant analytics. Technical issues include scaling, interface dependencies, in-memory computing, dealing with machine learning models as black boxes, and system stability. (Demirkan and Delen 2013)

During the fourth and final year students gain an ability to utilize more advanced machine learning models for both predictive and prescriptive analysis. They will learn to appreciate the inner workings of various learning methods and solve non-linear problems. The main focus of the studies will be on the computing side, i.e. that students learn how to create analytical systems. They will learn essential heuristic techniques and their mathematical explanation for problem solving, learning, and discovery.

Consequently the student should learn to predict events based on prior data through the use of machine learning. Machine learning models we refer to here are defined through the universal approximation theorem stating that any arbitrary continuous function can be estimated. This is done through a non-linear mapping of the input vector into a high-dimensional feature space, which in turn is connected to an output layer. (Haykin 2013) Students should thus learn to deal with high-dimensionality problems that allow them to master the implementation of advanced models for solving both regression and classification problems. This also requires the student to have a fundamental understanding of optimization techniques that can be used to demonstrate an optimal solution to a given problem.

Once students have an empirical understanding of dealing with machine learning models, focus is shifted towards creating services such as decision support systems and automated expert systems.

4.7 Thesis Module

In the thesis module the student learns to manage projects and understands how development is executed in agile projects. Students can express themselves in their native language both orally and in writing, as required by regulation. The student will be able to write a publication that summarizes the development of a project in a scientific manner.

5. Conclusions

The Information Technology field is rapidly developing and at the same time changing other fields it comes into contact with, everything from healthcare to the automotive industry. Rometty (2013) (CEO of IBM) commented the current technological shift towards analytical services as “...this is a thirty to fifty year, long-term project, which requires the next generation of computers, i.e. the self-learning computer”. During this time we will likely see the IT curriculum change many times and curricula for new disciplines develop, e.g. currently Business Analytics degree programs have hastily been developed at many Business Schools. The coming changes to the ACM IT curriculum recommendation should take into consideration Information Analytics in order for IT degree programs to stay relevant in the future. By tackling machine learning through development of analytical services and not through mathematics as is often done in Computer Science, we believe the area can be opened up to a greater engineering audience than before.

6. References

ACM: ACM Curricula Recommendations. 


AWS: AWS Elastic Beanstalk. 


Appendix

Details of our proposed IT curriculum for an Undergraduate Degree Program are shown in Figure 2 and Table 1. Figure 2 illustrates the curriculum structure at our university. We define the course units to be covered for each of the modules (with the exception for practical training) in Table 1.

<table>
<thead>
<tr>
<th>Module Title</th>
<th>Level</th>
<th>Course Units Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>General</td>
<td>Arcada 360; Introduction to Academic studies; Communications and Public Speaking; Nature of Code 1 - Introduction to Mathematical Programming; Statistics and Probability; Second Language</td>
</tr>
<tr>
<td>Web and Visualization</td>
<td>Basic</td>
<td>Web Development; Front-end Programming; Back-end Programming; Web services, Databases and CMS; Computer Architecture and Operating Systems; Nature of Code 2 - Vectors and Forces</td>
</tr>
<tr>
<td>Analytical Methods and Data Science</td>
<td>Professional</td>
<td>Information Visualization; Data Structures and Algorithms; IT-Law and Ethics; Concurrent Programming; Nature of Code 3 - Oscillation and Particle Systems; Network Protocols and Security;</td>
</tr>
<tr>
<td>Service Oriented Architectures and System Design</td>
<td>Professional</td>
<td>Network Communication and Cloud Technologies; Nature of Code 4 - Autonomous Agents and Cellular Automata; Nature of Code 5 - Fractals and Evolution of Code; Descriptive Analytics - Data/Text Mining; Software Defined Networks; Analysis and Design, UML and Design Patterns</td>
</tr>
<tr>
<td>Machine Learning and Decision Support System Development</td>
<td>Professional</td>
<td>Analytical System Design; Image and Speech Recognition Algorithms; Decision Support System Development and Verification; Predictive Analytics - Neural Networks; Prescriptive Analytics – Optimization; Process Optimization</td>
</tr>
<tr>
<td>Business Processes</td>
<td>Extension</td>
<td>Introduction to Business Administration; Introduction to Marketing; Digital Advertising; Introduction to Logistics; Intercultural Business; Introduction to Financial Management</td>
</tr>
<tr>
<td>Thesis Work</td>
<td>Thesis</td>
<td>Project Management; Academic Writing; Research Methodology and Seminar; Thesis Work (15 ECTS)</td>
</tr>
</tbody>
</table>

Table 1. Course Units Covered