

Erasmus Mundus Master Course

Information Technologies for Business Intelligence



Detailed Course Description

Academic Year 2014-2015

<p>University: Université Libre de Bruxelles (ULB) Department: Faculté des Sciences Appliquées Course ID: ADB (INFO-H-415) Course name: Advanced Databases Name and email address of the instructors: Esteban Zimányi (ezimanyi@ulb.ac.be) Web page of the course: http://cs.ulb.ac.be/public/teaching/infoh415 Semester: 1 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 24h. • Exercises: 24h. • Projects: 12h.
<p>Goals: Today, databases are moving away from typical management applications, and address new application areas. For this, databases must consider (1) recent developments in computer technology, as the object paradigm and distribution, and (2) management of new data types such as spatial or temporal data. This course introduces the concepts and techniques of some innovative database applications</p>
<p>Learning outcomes: At the end of the course students are able to</p> <ul style="list-style-type: none"> • Understand various different technologies related to database management system • Understand when to use these technologies according to the requirements of particular applications • Understand different alternative approaches proposed by extant database management systems for each of these technologies • Understand the optimization issues related to particular implementation of these technologies in extant database management systems.
<p>Readings and text books:</p> <ul style="list-style-type: none"> • R.T. Snodgrass, <i>Developing Time-Oriented Database Applications in SQL</i>, Morgan Kaufmann, 2000 • Jim Melton and Alan R. Simon, <i>SQL: 1999 - Understanding Relational Language Components</i>, Morgan Kaufmann, 2001 • Jim Melton, <i>Advanced SQL: 1999 - Understanding Object-Relational and Other Advanced Features</i>, Morgan Kaufmann, 2002 • Shashi Shekhar and Sanjay Chawla, <i>Spatial Databases: A Tour</i>, Prentice Hall, 2003.
<p>Prerequisites:</p> <ul style="list-style-type: none"> • Knowledge of the basic principles of database management, in particular SQL
<p>Table of contents:</p> <ul style="list-style-type: none"> • Active Databases Taxonomy of concepts. Applications of active databases: integrity maintenance, derived data, replication. Design of active databases: termination, confluence, determinism, modularisation. • Temporal Databases Temporal data and applications. Time ontology. Conceptual modeling of temporal aspects. Manipulation of temporal data with standard SQL. New temporal extensions in SQL 2011. • Object-Oriented and Object-Relational Databases Object-oriented model. Object Persistence. ODMG standard: Object Definition Language and Object Query Language. Object-relational model. Built-in constructed types. User-defined types. Typed tables. Type and table hierarchies. SQL standard and Oracle implementation. • Spatial Databases Application Domains of Geographical Information Systems (GIS), Common GIS data types and analysis. Conceptual Data Models for spatial databases. Logical data models for spatial databases: raster model (map algebra), vector model (OGIS/ SQL1999). Physical data models for spatial databases: Clustering methods (space filling curves), Storage methods (R-tree, Grid files).
<p>Assessment breakdown: 75% written examination, 25% project evaluation</p>

<p>University: Université Libre de Bruxelles (ULB) Department: Faculté des Sciences Appliquées Course ID: DBSA (INFO-H-417) Course name: Database Systems Architecture Name and email address of the instructors: Stijn Vansummeren (Stijn.vansummeren@ulb.ac.be) Web page of the course: http://cs.ulb.ac.be/public/teaching/infoh417 Semester: 1 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 24h. • Exercises: 12h. • Projects: 24h.
<p>Goals:</p> <p>In contrast to a typical introductory course in database systems where one learns to design and query relational databases, the goal of this course is to get a fundamental insight into the implementation aspects of database systems. In particular, we take a look under the hood of relational database management systems, with a focus on query and transaction processing. By having an in-depth understanding of the query-optimisation-and-execution pipeline, one becomes more proficient in administering DBMSs, and hand-optimising SQL queries for fast execution.</p>
<p>Learning outcomes:</p> <p>Upon successful completion of this course, the student:</p> <ul style="list-style-type: none"> • Understands the workflow by which a relational database management systems optimises and executes a query • Is capable of hand-optimising SQL queries for faster execution • Understands the I/O model of computation, and is capable of selecting and designing data structures and algorithms that are efficient in this model (both in the context of datababase systems, and in other contexts). • Understands the manner in which relational database management systems provide support for transaction processing, concurrency control, and fault tolerance
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom. <i>Database Systems: The Complete Book</i>, Prentice Hall, 2nd Edition, 2008. • Raghu Ramakrishnan and Johannes Gehrke. <i>Database Management Systems</i>. McGraw-Hill, 3rd Edition, 2002.
<p>Prerequisites:</p> <ul style="list-style-type: none"> • Introductory course on relational databases, including SQL and relational algebra • Course on algorithms and data structures • Knowledge of the Java programming language
<p>Table of contents:</p> <ul style="list-style-type: none"> • Query Processing With respect to query processing, we study the whole workflow of how a typical relational database management system optimises and executes SQL queries. This entails an in-depth study of: <ul style="list-style-type: none"> – translating the SQL query into a “logical query plan”; – optimising the logical query plan; – how each logical operator can be algorithmically implemented on the physical (disk) level, and how secondary-memory index structures can be used to speed up these algorithms; and – the translation of the logical query plan into a physical query plan using cost-based plan estimation. • Transaction Processing <ul style="list-style-type: none"> – Logging – Serializability – Concurrency control
<p>Assessment breakdown:</p> <p>70% written examination, 30% project evaluation</p>

<p>University: Université Libre de Bruxelles (ULB) Department: Faculté des Sciences Appliquées Course ID: DE (MATH-H-405) Course name: Applied Operational Research Name and email address of the instructors: Yves De Smet (yves.de.smet@ulb.ac.be) Web page of the course: http://code.ulb.ac.be/~yvdesmet/DEIT4BI Semester: 1 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 24h. • Exercises: 24h. • Projects: 12h.
<p>Goals: The goal of this course is to introduce some major chapters of operational research. The main aim is to illustrate how mathematical models and specific algorithms can be used to help decision makers facing complex problems (involving a large number of alternatives, multiple criteria, uncertain or risky outcomes, multiple decision makers, ...).</p>
<p>Learning outcomes: Upon successful completion of this course, the student:</p> <ul style="list-style-type: none"> • Is able to formulate and to solve basic decision problems; • Can identify the properties and limits of common decision models; • Is ready to deepen his/her knowledge in intermediate and advanced decision sciences courses.
<p>Readings and text books:</p> <ul style="list-style-type: none"> • C.D. Aliprantis, S.K. Chakrabarti, <i>Games and decision making</i>, Oxford University Press, 2000 • F.S. Hillier, G.J. Lieberman, <i>Introduction to Operations Research</i>, McGraw Hill, 2005 • H.A. Taha, <i>Operations Research: An introduction</i>, Pearson, Prentice Hall, 9th edition, 2010 • Ph. Vincke, <i>Multicriteria Decision-Aid</i>, J. Wiley, New York, 1992
<p>Prerequisites:</p> <ul style="list-style-type: none"> • Linear algebra • Basic course on algorithms • Probability and statistics
<p>Table of contents:</p> <ul style="list-style-type: none"> • Introduction to operational research The origin of operational research and decision sciences, some introductory examples. • Linear programming Models, graphical method, simplex algorithm, use of solvers. • Multicriteria Decision Aid Main concepts, introduction to multi-objective optimization, multi-attribute utility theory, outranking methods (ELECTRE & PROMETHEE), applications. • Decision under risk and uncertainty Common decision criteria: Maxmin, Maxmax, Hurwitz, Savage, Laplace. Expected utility. • Introduction to dynamic process (warehouse management, queuing systems)
<p>Assessment breakdown: 75% written examination, 25% project evaluation</p>

<p>University: Université Libre de Bruxelles (ULB) Department: Faculté des Sciences Appliquées Course ID: DW (INFO-H-419) Course name: Data Warehousing Name and email address of the instructors: Toon Calders (toon.calders@ulb.ac.be) Web page of the course: http://cs.ulb.ac.be/public/teaching/infoh419 Semester: 1 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 24h. • Exercises: 24h. • Projects: 12h.
<p>Goals:</p> <p>Relational and object-oriented databases are mainly suited for operational settings in which there are many small transactions querying and writing to the database. Consistency of the database (in the presence of potentially conflicting transactions) is of utmost importance. Much different is the situation in analytical processing where historical data is analyzed and aggregated in many different ways. Such queries differ significantly from the typical transactional queries in the relational model:</p> <ol style="list-style-type: none"> 1. Typically analytical queries touch a larger part of the database and last longer than the transactional queries; 2. Analytical queries involve aggregations (min, max, avg, ...) over large subgroups of the data; 3. When analyzing data it is convenient to see it as multi-dimensional. <p>For these reasons, data to be analyzed is typically collected into a data warehouse with Online Analytical Processing support. Online here refers to the fact that the answers to the queries should not take too long to be computed. Collecting the data is often referred to as Extract-Transform-Load (ELT). The data in the data warehouse needs to be organized in a way to enable the analytical queries to be executed efficiently. For the relational model star and snowflake schemes are popular designs. Next to OLAP on top of a relational database (ROLAP), also native OLAP solutions based on multidimensional structures (MOLAP) exist. In order to further improve query answering efficiency, some query results can already be materialized in the database, and new indexing techniques have been developed.</p> <p>The first and largest part of the course covers the traditional data warehousing techniques. The main concepts of multidimensional databases are illustrated using the SQL Server tools. The second part of the course consists of advanced topics such as data warehousing appliances, data stream processing, data mining, and spatial-temporal data warehousing. The coverage of these topics connects the data warehousing course with and serves as an introduction towards other related courses in the program. Several associated partners of the program contribute to the course in the form of invited lectures, case studies, and “proof of technology” sessions.</p>
<p>Learning outcomes:</p> <p>At the end of the course students are able to</p> <ul style="list-style-type: none"> • Understand the difference between operational databases and data warehouses • Understand the principles of multidimensional modeling • Understand the exploitation of a data warehouse for querying and reporting • Understand best practices and methodologies for data warehouse development • Understand the process of populating a data warehouse from internal and external sources
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Esteban Zimányi and Alejandro Vaisman. <i>Data Warehouse Systems: Design and Implementation</i>. Springer, 2014 • Matteo Golfarelli and Stefano Rizzi. <i>Data Warehouse Design: Modern Principles and Methodologies</i>. McGraw-Hill, 2009 • Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen. <i>Multidimensional Databases and Data Warehousing</i>. Morgan and Claypool Publishers, 2010 • Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker. <i>The Data Warehouse Lifecycle Toolkit</i>, 2nd ed. Wiley, 2008.
<p>Prerequisites:</p> <ul style="list-style-type: none"> • A first course on database systems covering the relational model, SQL, entity-relationship modelling, con-

- straints such as functional dependencies and referential integrity, primary keys, foreign keys.
- Data structures such as binary search trees, linked lists, multidimensional arrays.

Table of contents:

There is a mandatory project to be executed in three steps in groups of 3 students, using the tools learned during the practical sessions, being SQL Server, SSIS, SSAS, and SSRS. Below is the succinct summary of the theoretical part of the course:

- Foundations of multidimensional modelling
- Querying and reporting a multidimensional database with OLAP
- Methodological aspects for data warehouse development
- Populating a data warehouse: The ETL process

Assessment breakdown:

75% written examination, 25% project evaluation

<p>University: Université Libre de Bruxelles (ULB) Department: Faculté des Sciences Appliquées Course ID: BPM (INFO-H-420) Course name: Business Process Management Name and email address of the instructors: Toon Calders (toon.calders@ulb.ac.be) Web page of the course: http://cs.ulb.ac.be/public/teaching/infoh420 Semester: 1 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 24h. • Exercises: 24h. • Assignments and project: 12h.
<p>Goals:</p> <p>This course introduces basic concepts for modeling and implementing business processes using contemporary information technologies. The first part of the considers the <i>modeling</i> of business processes, including the control flow, and the data and resource perspectives. <i>Petri nets</i> will be used as a theoretical underpinning to formalize the different workflow patterns and unambiguously define the semantics of the different constructions in the workflow modelling languages. The workflow languages <i>Yet-another-workflow-Language</i> (YAWL) and the <i>Business Process Modelling and Notation</i> (BPMN) will be introduced in detail, as well as the main characteristics of the <i>Business Process Execution Language</i> (BPeL) for the composition of web services, and <i>Event-Driven Process Chains</i> (EPCs).</p> <p>The second part of the course then goes into the analysis, simulation, verification, and discovery of workflows. Static techniques to verify properties such as soundness and the option-to-complete at model level will be studied, as well as dynamic properties such as the compliance of an event log with respect to a given model. For the discovery of workflows, an overview of the main process mining techniques will be discussed.</p> <p>During the course the students have to perform a couple of modelling assignments in YAWL and BPMN. In the final project, students build a prototype system enacting one of the workflow modelled in their modelling assignments.</p> <p>Affiliated industrial partners of the Erasmus Mundus project will be involved in the course in the form of invited lectures, case studies, and “proof of technology” sessions. These lectures complement the academic coverage of the topic with a more business-oriented perspective and form a nice addition to provide a more complete picture of the Business Processing Modeling landscape.</p>
<p>Learning outcomes:</p> <p>At the end of the course students are able to</p> <ul style="list-style-type: none"> • Understand the value and benefit as well as the limitations of business process management • Understand the business process management life cycle • Model business processes in BPMN and YAWL • Construct a prototype business process in YAWL • Quickly master vendor-specific products in the BPM area
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Mathias Weske. <i>Business Process Management: Concepts, Languages, Architectures</i>. Springer, 2007 • Arthur H. M. ter Hofstede, Wil M. P. van der Aalst, Michael Adams, Nick Russell (Editors), <i>Modern Business Process Automation: YAWL and its Support Environment</i>. Springer, 2009. • Wil van der Aalst. <i>Process Mining: Discovery, Conformance and Enhancement of Business Processes</i>. Springer, 2012.
<p>Prerequisites:</p> <ul style="list-style-type: none"> • Basic programming skills: variables, control structures such as loops and if-then-else, procedures, object-oriented notions such as classes and objects, ... • Set theory (Notions such as set, set operations, sequence, multiset, function) and logics (mathematical notation and argumentation; basic proofs) • Basic graph theory (notions such as graphs, reachability, transitivity, ...) • Experience with modelling languages such as UML and ER diagrams is recommended.
<p>Table of contents:</p> <p>There is a mandatory project, split into several tasks during the whole period of the course offering, to be realized individually for the initial 3 assignments, and in group for the final assignment. The theoretical part</p>

of the course is dedicated to topics that allow the students to successfully carry out the project. Below is a high-level overview of the theoretical part of the course:

- Short overview of enterprise systems architecture and the place of business process management systems in it. The BPM life cycle.
- Modelling business processes: modelling the control flow, data and resource perspective.
- Enacting the business process models.
- Static and dynamic verification of process models; conformance checking.
- Discovering process models and other properties of processes through process mining.

Assessment breakdown:

50% oral examination, 50% project evaluation

<p>University: UFRT Department: Computer Science Dept. Course ID: ADW Course name: Advanced Data Warehousing Name and email address of the instructors: Dr. Patrick Marcel, Patrick.marcel@univ-tours.fr, Dr. Veronika Peralta, Veronika.Peralta@univ-tours.fr Web page of the course: Semester: 2 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 22 h. • Exercises: 16 h. • Project: 12 h.
<p>Goals:</p> <p>The aim of this course is to complement the course Data Warehouses (Semester 1) in its study of database technology used in Business Intelligence. A particular focus is given on the problems posed by heterogeneous data integration and data quality on the one hand, and on leveraging OLAP workload on the other hand. Classical notions of data warehousing and OLAP are recalled and developed: architecture, ETL, conceptual and logical design, query processing and optimization. Advanced topics like query personalization and recommendation are introduced.</p>
<p>Learning outcomes:</p> <p>Upon successful completion of this course, the student is able to:</p> <ul style="list-style-type: none"> • efficiently design, construct and query a data warehouse from real data sources, • define, measure and maintain data quality in the context of data warehousing.
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Ralph Kimball, Margy Ross, <i>The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling</i>, Second edition, John Wiley, 2002. • Matteo Golfarelli, Stefano Rizzi, <i>Data Warehouse Design: Modern Principles and Methodologies</i>, second edition, McGraw Hill, 2009. • Maurizio Lenzerini, Panos Vassiliadis, Matthias Jarke, Yannis Vassiliou, <i>Fundamentals of Data Warehouses</i>, Second edition, Springer, 2002. • Carlo Batini, Monica Scannapieco, <i>Data Quality Concepts, Methodologies and Techniques</i>, Springer, 2006.
<p>Prerequisites: Course DBSA and course DW (Semester 1).</p>
<p>Table of contents:</p> <ul style="list-style-type: none"> • Introduction <ul style="list-style-type: none"> – Data warehouse architecture and design – Advanced conceptual modeling • Quality <ul style="list-style-type: none"> – Quality models, diagnosis, correction, prevention • Loading <ul style="list-style-type: none"> – Integration, ETL • OLAP workload and querying <ul style="list-style-type: none"> – OLAP Workload, OLAP benchmarks – Formal models and languages, MDX, SQL99 • OLAP implementations <ul style="list-style-type: none"> – Query processing and optimization – In-memory, column-oriented storage • User-centric OLAP <ul style="list-style-type: none"> – Analytic sessions, query personalization and recommendations
<p>Assessment breakdown:</p> <p>Final written exam (50%), project (50%) assessed by oral presentations, demos, and written reports.</p>

<p>University: UFRT Department: Computer Science Dept. Course ID: BIS Course name: Business Intelligence Seminars Name and email address of the instructors: Patrick Marcel (Patrick.Marcel@univ-tours.fr) Web page of the course: Semester: 2 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 30 h. • Project: 20 h.
<p>Goals: Thanks to technological advances, the domain of business intelligence is witnessing today an increasing diversification and it addresses new application areas. This seminar covers current trends and recent developments in the domain of business intelligence. It also discusses the implications of business intelligence on individuals, organizations, and society in general. The seminar is designed and jointly taught by all consortium partners (whether full or associated partners, academic/research institutions or industrial companies), and will involve guest speakers presenting their organization, research topics, internships, and proposed Masters' thesis subjects for the second year of the master.</p>
<p>Learning outcomes: With this module, the student will 1) acquire a good understanding of the state of the art and the next evolutions and challenges in the domain, and 2) learn to synthesize, organize, and present scientific and technical information related to the domain of Business Intelligence, to make it clear to colleagues, and discuss it objectively.</p>
<p>Readings and text books: Research articles and white papers in the domain of Business Intelligence.</p>
<p>Prerequisites: All courses of Semester 1</p>
<p>Table of contents: The topics of the seminar vary from year to year, according to the guest speakers. In addition, students (in groups of maximum 5 persons) must realize a project supervised by one or more UFRT instructors, potentially under the co-supervision of one guest speaker. The topic is chosen in a list of topics defined jointly by UFRT instructors and invited speakers, and should be related to recent trends and developments in the domain of business intelligence. The group must write a report and make a presentation in front of their fellow students. The participation of students to presentations by guest speakers and fellow students is required.</p>
<p>Assessment breakdown: A project realized by students in group of 4 or 5, assessed by an oral presentation, a written report, and if applicable, a demo.</p>

<p>University: UFRT Department: Computer Science Dept. Course ID: IR Course name: Information Retrieval Name and email address of the instructors: Jean-Yves Antoine (Jean-Yves.Antoine@univ-tours.fr), Anaïs Lefeuvre (Anaïs.Lefeuvre@univ-tours.fr), Patrick Marcel (Patrick.Marcel@univ-tours.fr), Agata Savary (Agata.Savary@univ-tours.fr) Web page of the course: Semester: 2 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 20 h. • Exercises: 12 h. • Lab: 18 h.
<p>Goals:</p> <ul style="list-style-type: none"> • To study the problems posed by information retrieval • To study the processing, indexing, querying, organization and classification of textual documents • To study some basics of information ranking • To acquire a general idea of natural language applications and fundamentals • To acquire some practical skills in natural language applications related to information retrieval
<p>Learning outcomes: Upon successful completion of this course, the student is able to:</p> <ul style="list-style-type: none"> • To know how to index document corpus • To know how to design, construct and query a document database • To know how to personalize retrieval results • To have a basic understanding of linguistic modeling • To have a first experience on some NLP applications (name entities detection; textual information retrieval; text mining)
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, <i>Introduction to Information Retrieval</i>, Cambridge University Press. 2008. • James F. Allen, <i>Natural Language Understanding</i>. 2nd Edition, Benjamin Cummings, 1995. • Douglas Biber, <i>Variations across speech and writing</i>. Cambridge University Press, 1988. • Barbara J. Grosz, Karen Sparck-Jones, Bonnie Lynn Webber (Eds.), <i>Readings in Natural Language Processing</i>. Morgan Kaufmann Publ., 1986. • Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, <i>Spoken Language Processing: A Guide to Theory, Algorithm and System Development</i>. Prentice Hall, 2001. • Anand Rajaraman, Jure Leskovec and Jeffrey D. Ullman. <i>Mining of massive datasets</i>. http://i.stanford.edu/~ullman/mmds.html • Amy N. Langville, Carl D. Meyer, <i>Who's #1?: The Science of Rating and Ranking</i>. Princeton University Press, 2012
<p>Prerequisites: Courses DBSA and DW (Semester 1). Knowledge on automata and language theory is welcome.</p>
<p>Table of contents:</p> <ul style="list-style-type: none"> • Introduction: problems, retrieval processes, architecture, evaluation issues • Retrieval models: boolean model, vector space model, probabilistic model • Indexation • Introduction to ranking and rating • Ranking based on user interest: preferences, query expansion, recommendation • Markov ranking in Web search • Natural Language Processing for Information Retrieval: <ul style="list-style-type: none"> • General introduction : applications of NLP and linguistic level of description • Morphology : linguistic modeling (compound words), stemming, lemmatization • Terminology : motivation and applications • Morphology and syntax : POS tagging, named entities detection

- Syntax : parsing
- Semantic processing for information retrieval : latent semantic analysis

Assessment breakdown:

Final written exam (50%), project (50%) assessed by oral presentations, demos, and written reports.

<p>University: UFRT Department: Computer Science Dept. Course ID: KDDM Course name: Knowledge Discovery and Data Mining Name and email address of the instructors: Arnaud Giacometti (arnaud.giacometti@univ-tours.fr), Arnaud Soulet (arnaud.soulet@univ-tours.fr) and Haoyuan Li (haoyuan.li@univ-tours.fr) Web page of the course: Semester: 2 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 22 h. • Exercises: 16 h. • Lab: 12 h.
<p>Goals: The key objectives of this course are two-folds: (i) To give students a detailed understanding of the strengths and limitations of popular data mining techniques, and (ii) To understand the problems associated with the computational complexity issues in data mining.</p>
<p>Learning outcomes: Upon successful completion of this course, the student is able to:</p> <ul style="list-style-type: none"> • Prepare raw input data, and process it appropriately to provide suitable input for a wide range of data mining algorithms. • Understand the theoretical background of the main data mining algorithms. • Critically evaluate and select appropriate data mining algorithms. • Apply data mining algorithms, interpret and report the output appropriately. <p>More generally, students should be able to actively manage and participate in data mining projects executed by consultants or specialists in data mining</p>
<p>Readings and text books:</p> <ul style="list-style-type: none"> • J. Han and M. Kamber (2006), <i>Data Mining: Concepts and Techniques</i>, Morgan Kaufmann Publishers. • P.-N. Tan, M. Steinbach, V. Kumar (2005), <i>Introduction to Data Mining</i>, Addison-Wesley Publisher. • D. J. Hand, H. Mannila and P. Smyth (2001), <i>Principles of Data Mining</i>, The MIT Press.
<p>Prerequisites: Attendees must have prior knowledge on databases, algorithmic, data structures (trees, graphs). Some mathematical and statistical background will help.</p>
<p>Table of contents:</p> <ul style="list-style-type: none"> • Introduction to data mining: the Knowledge Discovery Process, Data Preparation • Classification Methods: <ul style="list-style-type: none"> – Basic Concepts and Decision Trees – Introduction to Artificial Neural Networks – Association Rule based Methods • Model Evaluation: Statistical Tests, ROC Analysis • Association Analysis: <ul style="list-style-type: none"> – Basic Concepts and Algorithms – Sequential Patterns, Data Streams • Clustering: <ul style="list-style-type: none"> – Basic Concepts and Algorithms – Partitioning and Hierarchical Methods • Bayesian Networks: Basic Concepts and Algorithms • Advanced Topics (seminars)
<p>Assessment breakdown: Project (40%) + written final examination (60%)</p>

<p>University: UFRT Department: Computer Science Dept. Course ID: XWT Course name: XML and web technologies Name and email address of the instructors: Dr Béatrice Bouchou Markhoff, (Beatrice.Bouchou@univ-tours.fr), Mirian Halfeld Ferrari Alves (mirian.halfeld@univ-tours.fr), Nizar Messai (Nizar.Messai@univ-tours.fr) Web page of the course: Semester: 2 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 22 h. • Exercises: 16 h. • Lab: 12 h.
<p>Goals:</p> <ul style="list-style-type: none"> • Understanding of the foundations of the web standard for data management, XML, with associated APIs, schema languages, query languages and transformation languages. • Basic knowledge on most important developments on the web, web services and semantic web.
<p>Learning outcomes: Upon successful completion of this course, the student is able to:</p> <ul style="list-style-type: none"> • Design XML documents • Design XML schemas (DTD, XML Schema) • Design integrity constraints for XML documents • Manage XML namespaces • Process XML documents w.r.t. schemas and integrity constraints (e.g. validation) • Query XML data with XPath and Xquery • Transform XML data with XSLT • Use Java APIs for XML processing • Discover how to express ontologies, create semantic annotations and express queries on semantic data (RDF, RDFS, OWL, SPARQL) • Discover how to design, modify and publish web services (SOAP, WSDL, UDDI)
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Serge Abiteboul, Ioana Manolescu, Philippe Rigaux, Marie-Christine Rousset and Pierre Senellart. <i>Web Data Management and Distribution</i>. Oxford Press, 2012. • Anders Møller and Michael I. Schwartzbach. <i>An Introduction to XML and Web Technologies</i>. Addison-Wesley, 2006.
<p>Prerequisites: Attendees must have prior knowledge on Algorithms and data structures (trees and graphs), language theory (finite state automata), First-order logics and AI</p>
<p>Table of contents:</p> <ul style="list-style-type: none"> • Introduction to semi-structured data and XML <ul style="list-style-type: none"> – XML document structure, infoset, namespace • Schema languages and validation process <ul style="list-style-type: none"> – DTD, XML Schema, (bottom-up unranked) tree automata • Navigating XML Trees and integrity constraints <ul style="list-style-type: none"> – XPath – Integrity constraints for XML • Querying (and transforming) XML documents <ul style="list-style-type: none"> – XQuery – XSLT • Programming with XML <ul style="list-style-type: none"> – API SAX – API DOM

- JAXP
- Introduction to Semantic Web technologies
 - Semantic annotations, RDF, SparQL
 - Concepts of ontology, RDF schema and OWL
- Introduction to Web Service Technologies
 - Service web description with SOAP and WSDL, and publication with UDDI

Assessment breakdown:

Written exam (60%) and Project (40%, oral presentation and written report)

<p>University: Ecole Centrale Paris (ECP) Department: Computer Science Department Course ID: VA Course name: Visual Analytics Name and email address of the instructors: Jean-Daniel Fekete (Jean-Daniel.Fekete@inria.fr) Web page of the course: (to be created) Semester: 3 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 24h • Laboratory: 24h • Project: 12h
<p>Goals: This course aims to help students understand the emerging, multidisciplinary field of VA, to familiarise them with current VA technology, and help them gain the foundations to building visual analytics tools and systems using real world data.</p>
<p>Learning outcomes: Upon completing the course, students will be able to:</p> <ul style="list-style-type: none"> • Understand basic concepts, theories and methodologies of Visual Analytics • Analyse data using appropriate visual thinking and visual analytics techniques • Present data using appropriate visual communication and graphical methods • Design and implement a Visual Analytics system for supporting decision making
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Edward Tufte, <i>Envisioning Information</i>, Graphics Press, 1990. • Robert Spence, <i>Information Visualisation: Design for Interaction</i>, Second Edition, Prentice Hall, 2007. • Colin Ware, <i>Information Visualisation: Perception for Design</i>, Second Edition, Morgan Kaufmann, 2004. <p>The course instructor will provide required weekly readings in the form of scientific articles, and will recommend further reading on each topic.</p>
<p>Prerequisites:</p> <ul style="list-style-type: none"> • Advanced Data Warehousing (ADW) • Knowledge Discovery and Data Mining (KD&DM)
<p>Table of contents:</p> <ul style="list-style-type: none"> • VA fundamentals: Theories, methodologies and techniques • Designing interactive graphics • Appropriate methods for different data types: Graphs, Hierarchies, Spatio-temporal data, High dimensional data • VA system design practices • Dashboard design
<p>Assessment breakdown: 30% class participation, 70% project (10% proposal, 20% intermediate, 40% final)</p>

<p>University: Ecole Centrale Paris (ECP) Department: Computer Science Department Course ID: CS&SW Course name: Corporate Semantics and Semantic Web Name and email address of the instructors: Gabriel Kepekian (gabriel.kepekian@atos.net) Web page of the course: (to be created) Semester: 3 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 24h • Laboratory: 24h • Project: 12h
<p>Goals: This course aims at presenting semantic technologies and the benefits to use them in companies for various applications such as semantic information search, recommendation, question and answering systems, search-based applications.</p>
<p>Learning outcomes:</p> <ul style="list-style-type: none"> • Provide the student with a deep understanding of the Semantic Web Technologies. • Ability to understand how to use Semantic Technologies for corporate application, with a special emphasise on the integration of unstructured content to enterprise structured data, • Be able to build an ontology and use it for a specific application.
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Pascal Hitzler, Markus Krotzsch, Sebastian Rudolph, <i>Foundations of Semantic Web Technologies</i>, Chapman & Hall/CRC, 2009. • Steffen Staab, Rudi Studer, <i>Handbook on Ontologies</i>, Springer, Second Edition, 2009. • John Davies, Rudi Studer, Paul Warren, <i>Semantic Web Technologies: Trends and Research in Ontology-based Systems</i>, Wiley, 2006 (published online) • David Taniar, Johanna Wenny Rahayu, <i>Web Semantics Ontology</i>, Idea Group Publishing, 2006.
<p>Prerequisites:</p> <ul style="list-style-type: none"> • XML and Web Technologies (X&WT) • Information Retrieval (IR)
<p>Table of contents:</p> <ul style="list-style-type: none"> • The Semantic Web Stack (RDF, RDFS, OWL, SKOS, SPARQL) • Ontology Learning and Life-Cycle • Linked Data • Adding Semantic to corporate data (Triple Store, Ontologies and OLAP, Ontologies and Databases) • Applications: Semantic Search, Question and Answering, Recommendation, Social Networks.
<p>Assessment breakdown: 50% written examination + 50% project evaluation</p>

<p>University: Ecole Centrale Paris (ECP) Department: Computer Science Department Course ID: DM&ML Course name: Data Mining and Machine Learning Name and email address of the instructors: Etienne Cuvelier (cuvelier.etienne@gmail.com), Antoine Cornuéjols (Antoine.Cornuejols@Iri.fr) Web page of the course: (to be created) Semester: 3 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 24h • Laboratory: 24h • Project: 27h
<p>Goals:</p> <p>The goals of this course are to allow the students to discover and practice advanced techniques of data mining and machine learning. Discover their principles, but also their variants, their applications and their weaknesses.</p>
<p>Learning outcomes:</p> <p>Upon successful completion of this course, the student will be able:</p> <ul style="list-style-type: none"> • to choose the best techniques to solve a given data mining or machine learning task, • to tune the parameters of the chosen technique, • to interpret the results of the chosen technique.
<p>Readings and text books:</p> <ul style="list-style-type: none"> • David J. Hand, Heikki Mannila and Padhraic Smyth, <i>Principles of Data Mining</i>, The MIT Press, 2001. • Trevor Hastie, Robert Tibshirani, Jerome Friedman, <i>The Elements of Statistical Learning: Data Mining, Inference, and Prediction</i>, Second Edition, Springer, 2009. • Luis Torgo, <i>Data Mining with R, learning with case studies</i>, Chapman & Hall/CRC, 2010.
<p>Prerequisites:</p> <ul style="list-style-type: none"> • Knowledge Discovery and Data Mining (KD&DM)
<p>Table of contents:</p> <ul style="list-style-type: none"> • Data Mining <ul style="list-style-type: none"> – Principal components analysis, factorial analysis. – Advanced clustering: spectral algorithms, galois lattices. – Regressions. – Analysis of non-tabular data types (graphs, symbolic, functional). • Problems and methods of Machine Learning <ul style="list-style-type: none"> – Formalization of the induction problem. Observations, hypothesis performance criterion. – Linear models. Generalization in kernel methods: SVM. – Deep neural networks. – Ensemble methods: boosting – Structured Data Learning: relational methods. – New problems: on-line learning, multi-task learning.
<p>Assessment breakdown:</p> <p>50% written examination + 50% project evaluation</p>

<p>University: Ecole Centrale Paris (ECP) Department: Computer Science Department Course ID: DM Course name: Decision Modelling Name and email address of the instructors: Valentina Ferretti (valentina.ferretti@polito.it) Web page of the course: (to be created) Semester: 3 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 24h • Laboratory: 24h • Project: 12h
<p>Goals: This course aims at presenting classical decision models with a special emphasis on decision making in uncertain situations, decision with multiple attribute, and decision with multiple stakeholders. During the course, various applications will be presented, emphasizing the practical interest and applicability of the models in real-world decision situations.</p>
<p>Learning outcomes:</p> <ul style="list-style-type: none"> • Provide the student with decision models and a better understanding of the validity of these decision models, • Ability to understand the three levels of decision analysis: representation of observed decision behaviour (descriptive analysis), decision aiding and recommendation (prescriptive analysis), and the design of artificial decision agents (normative analysis).
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Denis Bouyssou, Thierry Marchant, Marc Pirlot, Alexis Tsoukiás, Philippe Vincke, <i>Evaluation and decision models with multiple criteria: Stepping stones for the analyst</i>, Springer, International Series in Operations Research and Management Science Volume 86, 2006. • William W. Cooper, Lawrence M. Seiford, and Kaoru Tone, <i>Introduction to Data Envelopment Analysis and Its Uses</i>, Springer, 2006.
<p>Prerequisites:</p> <ul style="list-style-type: none"> • Decision Engineering (DE)
<p>Table of contents:</p> <ul style="list-style-type: none"> • Data envelopment Analysis: Analysis of the efficiency of production units • Decision under uncertainty, decision trees: theory, modeling and applications • Behavioural decision analysis: Empirical analysis of decision behaviour, cognitive decision biases, prospect theory • Outranking methods (theory and applications): Presentation of the Electre methods(Electre I, Electre 3, Electre Tri), reference based ranking. • Applications on a generic Decision platform: Decision Deck. Case studies and use of an open source platform for decision aid. • Group decision: Group decision, elicitation of a group decision model • Preference learning: Eliciting preference model for a decision maker, for several decision makers • Decision making using Multiple Objective Optimisation : Epsilon constraint method, applications, approximation algorithms, evolutionary algorithms, NSGA II
<p>Assessment breakdown: Assessment of the homework exercises (10%), Written exam (60%), Project (30%)</p>

<p>University: Ecole Centrale Paris (ECP) Department: Computer Science Department Course ID: II&R Course name: Introduction to Innovation and Research Name and email address of the instructors: Nacéra Bennacer (nacera.bennacer@supelec.fr) Web page of the course: (to be created) Semester: 3 Number of ECTS: 5</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 12h • Laboratory: 18h • Project: 30h
<p>Goals: The objectives of this course are to provide industrial and research presentations for students from the main BI software editors and clients as well as researchers in this domain. The European research context as well as intellectual properties, incubators and start-up creation will be presented. Students will also develop a research project in a collaborative way.</p>
<p>Learning outcomes:</p> <ul style="list-style-type: none"> • Provide the student with knowledge about intellectual properties, incubators and European research context (FP7 projects, ICT-labs, etc.) • Various seminars from key BI actors will be presented. • Ability to manage a research project for a client.
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Scientific papers will be distributed by the course lecturer according to the topics covered.
<p>Prerequisites:</p> <ul style="list-style-type: none"> • Business Intelligence Seminar (BIS)
<p>Table of contents:</p> <ul style="list-style-type: none"> • Seminars: innovation and research • Intellectual Property • Presentation of resources such as incubators • The European research context • Research project
<p>Assessment breakdown: Project evaluation 100%</p>

<p>University: Universitat Politècnica de Catalunya (UPC) Department: Department of Service and Information System Engineering Course ID: SOBI Course name: Service Oriented Business Intelligence Name and email address of the instructors: Alberto Abelló (aabello@essi.upc.edu) Web page of the course: http://www.fib.upc.edu/en/masters/miri/syllabus/SOBI-MIRI.htm Semester: 3 Number of ECTS: 6</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 36 h. • Problems and laboratories: 18 h. • Self-Study: 96 h.
<p>Goals:</p> <p>This course focuses on developing the student's skills to put their knowledge on service companies and benefiting from service facilities to build business intelligence solutions. On the one hand, we will analyze the specificity of this sector and present techniques to engineering their systems (i.e., Service Oriented Architecture). On the other hand, we will also analyze to which extent it is possible to consider Data Warehousing just a service and use these same techniques in its engineering methods.</p> <p>A joint real-world project potentially ready to be launched as a start up will be coordinated with other sibling subjects: BIP, SEAIT, SW and VBP. The analysis of the performance of technologies is the main outcome of SOBI, the management of the project will be developed in BIP, the interfaces in WS, the business plan in VBP, and the ethical part of the business model will be conducted in the humanities course (SEAIT).</p>
<p>Learning outcomes:</p> <p>Upon successful completion of this course, the student is able to:</p> <ul style="list-style-type: none"> • Knowledge <ul style="list-style-type: none"> – Understand the specificity of service companies. – Identify Business Intelligence as a service. – Recognise the characteristics and benefits of Infrastructure as a Service. – Recognise the characteristics and benefits of Software as a Service. – Recognise the characteristics and benefits of a Service Oriented Architecture (SOA). • Skills <ul style="list-style-type: none"> – Be able to develop Business Intelligence Service on Platform as a Service (specially on cloud databases). – Be able to work on Business Process as a Service and provide tools to analyse such processes (both qualitatively and quantitatively).
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Marie-Aude Aufaure and Esteban Zimányi, editors. <i>Proceedings of the 1st European Business Intelligence Summer School</i>, eBISS 2011, LNBIP 96. Springer, 2012. • Christian Baun. <i>Cloud computing : Web-based dynamic IT Services</i>. Springer, 2011. ISBN:978-3-642-20916-1. • Pramod Sadalage and Martin Fowler. <i>NoSQL distilled : a brief guide to the emerging world of polygot persistence</i>. Addison-Wesley, 2013. ISBN:978-0-321-82662-6. • Hasso Plattner and Alexander Zeier. <i>In-Memory Data Management</i>. Springer, 2011. ISBN:978-3-642-19362-0. • Marlo Dumas. <i>Fundamentals of business process management</i>. Spinger, 2011. ISBN:978-3-642-33142-8. <p>Other Literature:</p> <ul style="list-style-type: none"> • Tamer Özsu und Patrick Valduriez: <i>Principles of Distributed Database Systems</i>, Prentice Hall, third edition, 2011 • Anand Rajaraman, Jeffrey D. Ullman: <i>Mining of Massive Datasets</i>, Cambridge University. Press, 2012 • Will M. P. van der Aalst: <i>Process Mining: Discovery, Conformance and Enhancement of Business Processes</i>, Springer, 2011. • Scott E. Sampson. <i>Understanding Service Businesses: Applying Principles of Unified Services Theory</i>, 2nd Edition, John Wiley& Sons, 2001. • Thomas Erl. <i>Service Oriented Architecture</i>, Prentice Hall, 2006. • Mike Papazoglou et al., editors. <i>Service Research Challenges and Solutions for the Future Internet</i>, LNCS 6500, Springer 2010.

- David S. Linthicum. *Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide*, Addison-Wesley, 2009.

Prerequisites:

- Business Process Management (BPM)
- Advanced Data Warehousing (ADW)

Table of contents:

- Introduction to services: definition and specific characteristics
- Cloud Computing
 - IaaS
 - PaaS
 - SaaS
- Cloud databases
 - Distributed databases principles
 - CloudDB (BigTable)
 - MapReduce (Hadoop)
 - In-memory column stores (HANA)
 - Stream management
- BaaS
 - Service Oriented Architecture
 - ETL
- Service analysis
 - Qualitative
 - Quantitative

Assessment breakdown:

20% written examination, 40% problem classes and laboratories, 30% project, 10% peermarking

<p>University: Universitat Politècnica de Catalunya Department: Department of Service and Information System Engineering Course ID: Course name: Business Intelligence Project (BIP) Name and email address of the instructors: Oscar Romero (oromero@essi.upc.edu) Web page of the course: http://www.fib.upc.edu/en/masters/it4bi/syllabus/BIP-IT4BI.htm Semester: 3 Number of ECTS: 6</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 9 h. • Projects: 141 h.
<p>Goals:</p> <p>This course focuses on developing the student's skills to put their knowledge on software engineering and databases (as long as specific knowledge on project management introduced in this course) into practice, with the aim to develop information systems (IS) to support business intelligence (BI) processes within organizations. The course simulates an environment whose conditions are similar to those of a BI industrial project. So that, the students are required to work in a team, play roles, and successfully develop a project by gathering requirements from real stakeholders, planning the project, modeling the BI processes, analyzing and monitoring the project development to meet the end-user requirements, designing the system and incorporating automatic testing (in the form of Test-Driven Development and/or Behaviour-Driven Development), while documenting all the process. Eventually, a prototype is required. This course applies Agile Software Development techniques and the student will be introduced on how to adapt Agile for BI environments.</p> <p>A joint real-world project potentially ready to be launched as a start up will be coordinated with other sibling subjects: BIP, SEAIT, SW and VBP. The project management and implementation is the main outcome of BIP, the analysis of performance of the technologies will be developed in SOBI, the interfaces in WS, the business plan in VBP, and the ethical part of the business model will be conducted in the humanities course (SEAIT).</p>
<p>Learning outcomes:</p> <p>Upon successful completion of this course, the student is able to undertake BI projects and therefore:</p> <ul style="list-style-type: none"> • Lead and manage software projects for Business Intelligence, • Enable flexible and dynamic project environments with Agile Software Development principles and practices (when applicable), • Correctly identify and analyze the special needs of the project with regard to requirements, • Model and Deploy Business Intelligence Processes, • Design and deploy integrated data repositories, • Design and deploy data flow processes, • Assess the need of non-traditional tools / systems / methods for Big Data and Business Intelligence projects: NOSQL databases, integration techniques and data quality issues, • Incorporate automatic testing (BDD and/or TDD) to validate the system developed, • Incorporate effective data visualization techniques for Business Intelligence systems, • Successfully develop the project in a well-rounded, disciplined and methodological manner, • Strength the student's team work skills, such as the ability to reach agreements, play a specific team role and reuse and continue other teammate's work, • Perform oral presentations and defense in real business environments with customers.
<p>Readings and text books:</p> <ul style="list-style-type: none"> • M. Golfarelli, S. Rizzi, <i>Data Warehouse Design</i>, McGraw Hill, 2009 • A. Fox, D. Patterson, <i>Engineering software as a service: an Agile approach using Cloud Computing</i>, Strawberry Canyon, 2012. • B. Meyer, <i>Agile! The Good, The Hype and the Ugly</i>, Springer, 2014.
<p>Prerequisites:</p> <ul style="list-style-type: none"> • Advanced Databases (AD) • Business Process Management (BPM) • Advanced Data Warehousing (ADW) • Information Retrieval (IR) • Knowledge Discovery and Data Mining (KDDM)

Table of contents: At the beginning of the course a real case study is presented to the students. They will need to develop in the shape of a project course. During the first weeks there will be an introduction to key concepts prior to start developing it:

- Specificities of BI projects
- Agile project management techniques for BI systems.
 - Requirement elicitation for BI (Agile approach)
 - Project management techniques: SCRUM, KANBAN
 - Automatic testing: BDD, TDD

From there on, the students are expected to develop the whole project until completion throughout 6 well-defined stages:

- Formal specification and analysis
- Modeling BI processes
- Design (including choose the appropriate service-oriented architecture for the system)
- Project management
- Validation / testing
- Project defense

The project is expected to be developed in the course sessions (under supervision of the teacher), and as teamwork (with no supervision). The course sessions (following the Agile Sprint concept) follow a project based learning approach (PBL) in which the students are required to play a specific role within their team, undertake team discussions, reach agreements and champion their decisions.

Assessment breakdown:

30% project sprints, 30% project management, 20% project defense, 20% peer-marking (this mark is assigned by the lecturer based on the mark given by your teammates)

<p>University: Universitat Politècnica de Catalunya (UPC) Department: Department of Service and Information System Engineering Course ID: Course name: Web Services (WS) Name and email address of the instructors: Carles Farré (farre@essi.upc.edu) Web page of the course: http://www.fib.upc.edu/en/masters/miri/syllabus/WS-MIRI.htm Semester: 3 Number of ECTS: 6</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 26 h. • Lab sessions: 28 h. • Autonomous work: 81 h. • Exam preparation: 15 h.
<p>Goals:</p> <p>This course focuses on the need for systems interoperability and how Web services, an umbrella concept for different cross-platform solutions based on Web standards, attempt to overcome the many challenges that distributed information systems have addressed in various ways (but not always successfully) in the past. By the end of the course, students will have learned the relevant concepts related to the nature, characteristics and types of Web Services and acquired some experience in consuming, designing, constructing and maintaining services located on the Web.</p> <p>A joint real-world project potentially ready to be launched as a start up will be coordinated with other sibling subjects: BIP, SEAIT, SOBI and VBP. The ability to create Web-based interfaces for the project will be the main outcome of WS, the analysis of the performance of the required technologies will be dealt with at SOBI, the management of the project will be developed at BIP, the business plan in VBP, and the ethical part of the business model will be conducted in the humanities course (SEAIT).</p>
<p>Learning outcomes:</p> <p>Upon successful completion of this course, the student is able to:</p> <ul style="list-style-type: none"> • Understand the fundamental Web technologies that are the basis for the development of web services, • Know the different protocols and communication standards for web services, • Design and implement software that interacts with web services and public or private web APIs, • Design and implement web services, selecting and using the technologies and tools that are most appropriate in each case, • Test and monitor web services, selecting and using the technologies and tools that are most appropriate in each case.
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Gustavo Alonso, Fabio Casati, Harumi Kuno, Vijay Machiraju. <i>Web Services. Concepts, Architectures and Applications</i>, Springer, 2004. • Robert Daigneau. <i>Service Design Patterns: Fundamental Design Solutions for SOAP/WSDL and RESTful Web Services</i>. Addison-Wesley Professional, 2011. • Thomas Erl, Benjamin Carlyle, Cesare Pautasso, Raj Balasubramanian. <i>SOA with REST: Principles, Patterns & Constraints for Building Enterprise Solutions with REST</i>, Prentice Hall, 2012. • Michael P. Papazoglou. <i>Web Services: Principles and Technology</i>, Prentice Hall, 2008. • Leon Shklar, Rich Rosen. <i>Web Application Architecture: Principles, Protocols and Practices</i>. Second Edition, John Wiley & Sons, 2009.
<p>Prerequisites:</p> <ul style="list-style-type: none"> • XML and Web Technologies (X&WT)
<p>Table of contents:</p> <ul style="list-style-type: none"> • Introduction • Origins & Precedents: Fundamentals of Distributed Systems. Middleware, SOA • Core Web Technologies: <ul style="list-style-type: none"> – The Fundamentals: URIs. HTTP. Proxies, caches, cookies – Browser-Based Computing: JavaScript, DOM, AJAX – Server-Side Computing: CGI, PHP, Java Servlets – Web Data Exchange Formats: XML, JSON

- Core WS Protocols
 - SOAP and WSDL
 - RESTful WS
- WS Development
 - Properties of a service development methodology
 - Qualities of service development methodology
 - Web services development lifecycle
 - Service analysis, design and construction
 - Design Patterns for Web Service Development
- Securing WS
 - General Concepts
 - Securing RESTful Web Services
 - XML Security Standards
 - Securing WS-* Web Services
- Advanced Topics

Assessment breakdown:

30% Written final exam, 30% Weekly lab Sessions, 20% Lab project, 20% Presentations

The instructors will present the course contents using slides or some other material.

There will be also lecture sessions in which students will be required to prepare on their own and present to the class a certain topic.

During the lab, during the first 9 weeks of the course, and after a brief introduction, the students will be required to complete a certain number of tasks using the computer in accordance with a work plan.

During the last 5 lab sessions of the course, the students, in groups of 3-4, will design and implement a web service project that will be coordinated with the other sibling subjects: BIP, SEAIT, SOBI and VBP. The lab sessions will be used to work in the project, discuss and fix doubts and problems, plan and manage goals, and present results.

Some course contents are not presented in the lectures and must be studied by the students on their own. Teachers will indicate which contents should be studied and the teaching resources that might be employed.

<p>University: Universitat Politècnica de Catalunya (UPC) Department: Department of Management Course ID: Course name: Viability of Business Projects (VBP) Name and email address of the instructors: Marc Eguiguren (marc.eguiguren@upc.edu) Web page of the course: http://www.fib.upc.edu/en/masters/it4bi/syllabus/VBP-IT4BI.htm Semester: 3 Number of ECTS: 6</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 36 h. • Projects in the classroom: 18 h. • Projects: 56 h. • Self-Study: 40 h.
<p>Goals:</p> <p>University graduates can find themselves in the situation of having to analyse or take on the project of starting their own business. This is especially true in the case of computer scientists in any field related to Business Intelligence (BI) or more generally, in the world of services. There are moments in one's professional career at which one must be able to assess or judge the suitability of business ventures undertaken or promoted by third parties, or understand the possibilities a new service has for success. It is for this reason that this subject focuses on providing students with an understanding of the main techniques used in analysing the viability of new business ventures: business start-up or the implementation of new projects in the world of services or in the specific field of BI. This project-oriented, eminently practical subject is aimed at each student's being able to draft as realistic a business plan as possible.</p> <p>A joint real-world project potentially ready to be launched as a start up will be coordinated with other sibling subjects: BIP, SEAIT, SW and SOBI. The creation of the business idea, the business model and the development of the business plan are the main outcomes of VBP, the analysis of the performance of the required technologies will be dealt with at SOBI, the management of the project will be developed at BIP, the interfaces at WS, and the ethical part of the business model will be conducted in the humanities course (SEAIT).</p>
<p>Learning outcomes:</p> <p>Upon successful completion of this course, the student is able to:</p> <ul style="list-style-type: none"> • Being able to analyze the external situation to determine business innovative ideas in the field of BI • Around an innovative BI project, being able to build a reasonable and ethically solid business plan • Building a solid and convincing speech about a business idea and a business plan • Training the students to build a P&L forecast and a forecasted treasury plan for a starting company • Understanding and being able to apply the different instruments to finance the company, both debt instruments or private equity and venture capital sources • Understand and appreciate the role of the entrepreneur in modern society
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Rhonda Abrams, Eugène Kleiner. <i>The Successful Business Plan</i>. The Planning Shop, 2003. • Rob Cosgrove. <i>Online Backup Guide for Service Providers</i>, Cosgrove, Rob, 2010. • Peter Drucker. <i>Innovation and Entrepreneurs</i>. Butterworth-Heinemann, Classic Drucker Collection edition, 2007. • Robert D. Hisrich, Michael P. Peter, Dean A. Shepherd. <i>Entrepreneurship</i>. Mc Graw Hill, 6th Ed., 2005. • Mike McKeever. <i>How to Write a Business Plan</i>. Nolo, 2010. • Lawrence W. Tuller. <i>Finance for Non-Financial Managers and Small Business Owners</i>. Adams Business, 2008. <p>Other Literature:</p> <ul style="list-style-type: none"> • M. Eguiguren, E. Barroso. <i>Empresa 3.0: políticas y valores corporativos en una cultura empresarial sostenible</i>. Pirámide, 2011. • M. Eguiguren, E. Barroso. <i>Por qué fracasan las organizaciones: de los errores también se aprende</i>. Pirámide, 2013.
<p>Prerequisites:</p> <p>Having some previous knowledge or experience in business administration is an additional asset.</p>

Table of contents: This course focuses on developing a BI or services oriented business plan. So that, the students are expected to reuse and consolidate any previous knowledge on databases, software engineering and BI obtained in previous courses to develop a comprehensible, sustainable and profitable business.

The course is structured in 14 well-defined stages:

- Introduction to the course and key aspects of a business idea
- The entrepreneur's role in society, characteristics and profile
- Innovation and benchmarking Axis 1) Identification of long-term market megatrends
- Innovation and benchmarking axis 2) Technological evolution as a source of ideas. Technology applied to industry.
- Axis of innovation and benchmarking 3) ethical business models as a source of innovation and ideas
- From the idea to the company. Contents of the business plan. Market research.
- Competitive advantages. SWOT Analysis
- Marketing plan: strategic marketing, distribution and product
- Marketing plan: price and promotion strategies
- The human team in a small innovative company
- Different kind of societies. Fiscal basics for entrepreneurs
- Need of resources. Building the balance sheet at the beginning of the company
- Building a forecasted P&L for the first two years. Cash-Flow
- Revising the initial balance sheet and building the forecasted balance sheet for year one
- Treasury plan, Identifying long and short term financial needs
- Conventional long and short term financial instruments
- Private equity: founders, fools, friends & family, venture capital. Their limitations. Cautions to be taken and how they work.
- Presenting the plan to possible simulated or real investors

The business plan is expected to be partially developed in internal activities (under supervision of the teacher), and in external activities, always as teamwork (with no supervision).

Assessment breakdown:

The assessment is based on student presentations and the defence of the business plan before a jury comprising course faculty members and - optionally - another member of the teaching staff or guest professional.

Throughout the course there will be four evaluative milestones:

- presentation of the innovative business model,
- presentation of the marketing plan,
- presentation of the business plan as a whole, that will include an evaluation about ethics and sustainability of the project together with SEAIT,
- analysis of the financial plan and the proposal to investors.

The presentation simulates a professional setting. Accordingly, the following aspects will also be assessed: dress, formal, well-structured communication, etc.

In order to be able to publicly defend the business plan, students must have attended at least 70% of the classes and teams must have delivered on time the activities that have been planned during the course. The plan is the result of teamwork, which will be reflected in the grade given to the group as a whole. Each member of the group will be responsible for part of the project and will be graded individually on his or her contribution.

This approach is designed to foster teamwork, in which members share responsibility for attaining a common objective.

<p>University: Technische Universität Berlin (TUB) Department: School of Electrical Engineering and Computer Science Course ID: BDASEM Course name: Big Data Analytics Seminar Name and email address of the instructors: Volker Markl (sekr@dima.cs.tu-berlin.de) Web page of the course: http://www.dima.cs.tu-berlin.de Semester: 3 Number of ECTS: 3</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Lectures: 30h. • Exercises: 30h.
<p>Goals: Participants of this seminar will acquire knowledge about recent research results and trends in the analysis of web-scale data. Through the work in this seminar, students will learn the comprehensive preparation and presentation of a research topic in this field. In order to achieve this, students will get to read and categorise a scientific paper, conduct background literature research and present as well as discuss their findings.</p>
<p>Learning outcomes: After the course, students will be able to critically read and evaluate scientific publications, and to conduct background research. They will be capable of preparing for and giving oral presentations on research topics for an expert audience, of analyzing the state of the art of a research topic, and of summarizing it in a scientific paper. They should also understand techniques used in the scientific community like peer reviews, conference presentations, and defenses of the findings after their presentation, as well as they should understand methods for large-scale data analytics.</p>
<p>Readings and text books: At the beginning of the semester students will receive a set of primary literature, which consists of a basic item for every participant. Then students will learn about presentation techniques and guidelines on how to read scientific papers. This is be extended by learning how to write texts specially in the context of the English language. Students should use secondary sources to research the topic assigned to them in the seminar, which should go beyond the supplied primary literature. Next to conventional sources like the internet students are required to use research journals and articles published at information management conferences such as WWW, VLDB, or SIGMOD.</p> <ul style="list-style-type: none"> • Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom: <i>Database Systems: The Complete Book</i>, Prentice Hall, second edition, 2009. • Tom White. <i>Hadoop: The Definitive Guide</i>, Third Edition. O'Reilly Media, 2012. • Jimmy Lin, Chris Dyer <i>Data-Intensive Text Processing with MapReduce</i>. Morgan & Claypool, 2010.
<p>Prerequisites:</p> <ul style="list-style-type: none"> • Database Systems Architecture (DBSA)
<p>Table of contents: Both the sciences and industry are currently undergoing a profound transformation: large-scale, diverse data sets - derived from sensors, the web, or via crowd sourcing - present a huge opportunity for data-driven decision making. This data poses new challenges in a variety of dimensions: in its unprecedented volume, in the speed at which it is generated (its velocity) and in the variety of data sources that need to be integrated. A whole new breed of systems and paradigms is currently developed to be able to cope with that these challenges. The field of Big Data Analytics deals with the technological means of gaining insights from huge amounts of data. In this seminar, students will review the current state of the art in this field.</p>
<p>Assessment breakdown: The grade of the module will be composed from the results of the presentation (50%) and the written seminar report (50%) for this presentation.</p>

<p>University: Technische Universität Berlin (TUB) Department: School of Electrical Engineering and Computer Science Course ID: : IDB-PRA (0434 L 468) Course name: Implementation of a Database Engine (practice project) Name and email address of the instructors: Volker Markl (sekr@dima.cs.tu-berlin.de) Web page of the course: http://www.dima.cs.tu-berlin.de Semester: 3 Number of ECTS: 6</p>
<p>Course breakdown and hours: 4 hrs. plenary sessions per week, 8 hrs. home and project work, totals 180 hrs./term</p> <ul style="list-style-type: none"> • Plenary sessions: 60 hrs. • Project work/Preparation & consolidation: 120 hrs.
<p>Goals: The global data volume is increasing dramatically each year. Understanding how to store, process and manage these huge amounts of data efficiently is a key requirement for software engineers and data analysts in the modern IT world. This project (following the corresponding lecture topics of IDB_VL – Database Internals & Scalable Data Processing) will teach students both the fundamentals of data processing in traditional single-node database systems and how to scale out these techniques to huge amounts of data in large-scale, distributed environments.</p>
<p>Learning outcomes: Upon successful completion of this course, the participants of this course/project will achieve in-depth knowledge about both the fundamentals of data processing in traditional single-node database systems and how to scale out these techniques to huge amounts of data in large-scale, distributed environments. During the implementation project, students will get hands-on experience with important data processing techniques by implementing several components of a relational database system and by using parallel programming platforms like Apache Hadoop or Nephele/PACT.</p> <p>The course is principally designed to impart: Technical skills 30%; Method skills 30%, System skills 30%, social competence: 10%.</p>
<p>Readings and text books: Primary literature</p> <ul style="list-style-type: none"> • Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom: <i>Database Systems: The Complete Book</i>, Prentice Hall, second edition, 2009. <p>Other literature</p> <ul style="list-style-type: none"> • R. Elmasri, S.B. Navathe: <i>Fundamentals of Database Systems</i>, Sixth edition, Addison-Wesley, 2010 • J. Gray, A. Reuter: <i>Transaction Processing: Concepts and Techniques</i>, Morgan Kaufman, 1992 • T. Özsu, P. Valduriez: <i>Principles of Distributed Database Systems</i>, Second edition, Springer, 2011 • K.-U. Sattler, A. Heuer, G. Saake: <i>Datenbanken - Konzepte und Sprachen</i>, Fifth edition, mitp Verlag, 2013 • T. Härder, E. Rahm: <i>Datenbanksysteme: Konzepte und Techniken der Implementierung</i>, Springer, Second edition, 2001 • A. Kemper, A. Eickler: <i>Datenbanksysteme: Eine Einführung</i>, Oldenburg, Ninth Auflage 2013
<p>Prerequisites: A basic course on database systems. Knowledge of data modeling, relational algebra, and SQL as well as a very good command of Java, or possibly C/C++/C#, programming is required to participate in the course.</p>
<p>Table of contents: During the implementation project, students will implement several components of a relational database system, using parallel programming platforms like Apache Hadoop or Nephele/PACT: Table Page Layout, ARC, Buffer Pool Manager, B-Tree, Operators, Optimizer, MapReduce</p>
<p>Assessment breakdown:</p> <ul style="list-style-type: none"> • Presentation/demo of the implementation project 30% • Successful completion of the implementation project 70%

<p>University: Technische Universität Berlin (TUB) Department: School of Electrical Engineering and Computer Science Course ID: AIM1 - HDIS (0434 L 440) Course name: Heterogeneous and Distributed Information Systems Name and email address of the instructors: Ralf Kutsche (sekr@dima.cs.tu-berlin.de) Web page of the course: http://www.dima.cs.tu-berlin.de Semester: 3 Number of ECTS: 6</p>
<p>Course breakdown and hours: Integrated course: 4 hrs. plenary sessions per week, 8 hrs. home and project work, totals 180 hrs./term</p> <ul style="list-style-type: none"> • Lectures & Labs (integrated): 60 hrs. • Home/Project Work: 120 hrs.
<p>Goals: This course addresses master students with a focus on database systems, information management and business intelligence, in order to achieve deep knowledge in modern distributed heterogeneous information infrastructures. Heterogeneity in data models, distributed data organization/software architectures as well as interoperability and middleware platforms for HDIS (FIS, P2P, CS, ...), and, finally their persistency concepts and services will be studied in deep detail. Additional support by semantic concepts (ontologies, ...), model/metamodel and metadata management guides the way towards model-based development of HDIS and their applications in industry and public services.</p>
<p>Learning outcomes: Upon successful completion of this course, the participants of this course will achieve deep conceptual, methodical, technical and practical knowledge in requirements analysis, design, architecture and development of heterogeneous and distributed information systems. This includes firstly classical knowledge about federated databases and mediator-based information systems (tight or loose coupling wrt. the dimensions of distribution, heterogeneity and autonomy). Secondly, the students can handle different paradigms of heterogeneous information infrastructures and their management (e.g., FedIS, CS, P2P) and interoperability architectures ("middleware"). Finally, modern model-based concepts for the development, integration and evolution of arbitrary information infrastructures, and –under this conceptual frame– model, metamodel, and metadata management as well as semantic concepts will be brought into practical experience by a larger seminar project work.</p> <p>The course is principally designed to impart: technical skills 50%, method skills 30%, system skills 10%, social skills 10%</p>
<p>Readings and text books: For each topic during this course appropriate text books, journal and conference papers, PhD dissertations, technical reports, and industrial standards will be used.</p>
<p>Prerequisites:</p> <ul style="list-style-type: none"> • Database Systems Architecture (DBSA) • A course on Software Engineering • A course on Advanced Information Modelling (e.g., INFMOD at TUB), strongly recommended
<p>Table of contents:</p> <ul style="list-style-type: none"> • Foundations/Terminology of HDIS (FDBS, FIS, MBIS) • Dimensions of HDIS: Distribution, Heterogeneity, Autonomy • Heterogeneous Data Models in HDIS: structured, semistructured, unstructured • Distributed Data Organisation and Software Architectures of HDIS (FIS, P2P, CS, ...) • Interoperability and Middleware Platforms for HDIS • Persistency Services • Semantic Concepts in HDIS • Metadata Standards and Management in HDIS • Model-based Development of HDIS • Applications from Industry and Public Services
<p>Assessment breakdown: Seminar Talk 25%, Technical Report 40%, Project 35%, Homework & Lab Assessment 5%</p>

<p>University: Technische Universität Berlin (TUB) Department: School of Electrical Engineering and Computer Science Course ID: IMPRO3 (0434 L 483) Course name: Big Data Analytics Projects Name and email address of the instructors: Volker Markl (sekr@dima.cs.tu-berlin.de) Web page of the course: http://www.dima.cs.tu-berlin.de Semester: 3 Number of ECTS: 9</p>
<p>Course breakdown and hours:</p> <ul style="list-style-type: none"> • Projects: 60 h
<p>Goals:</p> <p>In this course you will learn to systematically analyze a current issue in the information management area and to develop and implement a problem-oriented solution as part of a team. You will learn to cooperate as team member and to contribute to project organization, quality assurance and documentation. The quality of your solution has to be proven through analysis, systematic experiments and test cases. Examples of IMPRO projects carried out in recent semesters are a tool used to analyse Web 2.0 Forum data, an online multiplayer game for mobile phones, implementation and analysis of new join methods for a cloud computing platform or the development of data mining operations on the massively parallel system Hadoop as part of the Apache open source project Mahout.</p>
<p>Learning outcomes:</p> <p>After the course, students will be able to understand methods for large-scale data analytics and to solve large-scale data analytics problems. They will be capable of designing and implementing large-scale data analytics solutions in a collaborative team.</p>
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom: <i>Database Systems: The Complete Book</i>, Prentice Hall, second edition, 2009. • Anand Rajaraman, Jeffrey D. Ullman, <i>Mining of Massive Datasets</i>, Cambridge 2010.
<p>Prerequisites:</p> <ul style="list-style-type: none"> • Heterogeneous and Distributed Information Systems (H&DIS)
<p>Table of contents:</p> <p>Both the sciences and industry are currently undergoing a profound transformation: large-scale, diverse data sets - derived from sensors, the web, or via crowd sourcing - present a huge opportunity for data-driven decision making. This data poses new challenges in a variety of dimensions: in its unprecedented volume, in the speed at which it is generated (its velocity) and in the variety of data sources that need to be integrated. A whole new breed of systems and paradigms is currently developed to be able to cope with that these challenges.</p> <p>The field of Big Data Analytics deals with the technological means of gaining insights from huge amounts of data. Students will conduct projects that deal with applying data mining algorithms to large datasets. For that, students will learn to use so called Parallel Processing Platforms, systems that execute parallel computations with terabytes of data on clusters of up to several thousand machines.</p> <p>At the start of the project, a student will receive a topic as well as some information material. The team, with the assistance of the lecturer, will decide on a project environment with the suitable tools for team work, project communication, development and testing. Next, the problem will have to be analyzed, modelled and decomposed into individual components, from which tasks are derived that are subsequently assigned to smaller teams or individuals. At weekly project meetings, the project team presents progress and milestones that have been reached. In consultation with the lecturer, it is decided which further steps to take. The project is concluded with a final report, a project poster as well as a final presentation which includes a demonstration of the prototype.</p>
<p>Assessment breakdown:</p> <p>The overall grade for the module consists of the results of ‘exam equivalent’ course work (Prüfungsäquivalente Studienleistungen Päs). The following are included in the final grade:</p> <ul style="list-style-type: none"> • Active participation in the project (10%) • Prototype with test cases (50%) • Documentation (10%) • Final Report (10%)

- Project Poster (10%)
- Final presentation (10%)

<p>University: Technische Universität Berlin (TUB) Department: School of Electrical Engineering and Computer Science Course ID: AIM3 - SDADM (0434 L 472) Course name: Scalable Data Analysis and Data Mining Name and email address of the instructors: Volker Markl (sekr@dima.cs.tu-berlin.de) Web page of the course: http://www.dima.cs.tu-berlin.de Semester: 3 Number of ECTS: 6</p>
<p>Course breakdown and hours: Integrated course: 4 hrs. plenary sessions per week, 8 hrs. home and project work, totals 180 hrs./term</p> <ul style="list-style-type: none"> • Plenary sessions: 60 hrs. • Exercises/practice: 30 hrs. • Home/Project Work: 60 hrs. • Preparation & Consolidation: 30 hrs.
<p>Goals: The focus of this module is to get familiar with different parallel processing platforms and paradigms and to understand their feasibility for different kinds of data mining problems. For that students will learn how to adapt popular data mining and standard machine learning algorithms such as: Naïve Bayes, K-Means clustering or PageRank to scalable processing paradigms. And subsequently gain practical experience in how to implement them on parallel processing platforms such as Apache Hadoop, Stratosphere/Apache Flink and Apache Giraph.</p>
<p>Learning outcomes: Recent advances in technology have led to rapid growth of big data. This led to the need for cost efficient and scalable analysis algorithms. In this course concepts for scalable analysis of big data sets will be presented and applied using open source technologies. Participants of this module will gain an in-depth understanding of concepts and methods as well as practical experience in the area of scalable data analysis and data mining. The course is principally designed to impart: technical skills 50%, method skills 30%, system skills 10%, social skills 10%</p>
<p>Readings and text books:</p> <ul style="list-style-type: none"> • Anand Rajaraman, Jeffrey David Ullman. <i>Mining of Massive Datasets</i> (Free online: http://infolab.stanford.edu/~ullman/mmds/book.pdf) • Ian H. Witten, Eibe Frank, Mark A. Hall. <i>Data Mining: Practical Machine Learning Tools and Techniques</i>, Third Edition. Morgan Kaufmann, 2011. • Tom White. <i>Hadoop: The Definitive Guide</i>, Third Edition. O'Reilly Media, 2012.
<p>Prerequisites: Basic course on Database Systems as well as good Java programming skills are required. A basic understanding of Probability and Statistics as well as Linear Algebra is helpful.</p>
<p>Table of contents:</p> <ul style="list-style-type: none"> • MapReduce + HDFS / MapReduce (in detail) + joins • Apache Flink / Spark • Review linear algebra, probability, statistics • Introduction to data mining (including data preparation) • Classification (including feature extraction) • Clustering • Dimensionality reduction (including feature extraction) • Collaborative filtering • Streaming • Network analysis • Statistical NLP • Privacy, legal issues • Visualization analytics
<p>Assessment breakdown: Written homework (20 %), protocolled practical project work (50 %), oral feedback session (30 %)</p>