

1.1 Organizzazione e contenuti dei moduli

I moduli A e B saranno svolti in aula ed in laboratorio e prevedono la frequenza degli allievi per cinque giorni settimanali per complessive 5 ore giornaliere, impiegate in lezioni in aula, attività di laboratorio e realizzazione di progetti didattici.

Il Modulo C prevede lo svolgimento delle attività di *training on the job* che impegneranno gli allievi per i restanti 6 mesi, per 5 (cinque) giorni a settimana, dal lunedì al venerdì, con un impegno quotidiano di 8 ore. Durante il *training on the job* gli allievi metteranno in pratica quanto appreso in aula attraverso lo sviluppo e la realizzazione di progetti che saranno oggetto di valutazione finale, eventualmente, affiancando i ricercatori impegnati nella realizzazione del Progetto di ricerca.

1.2 Modulo A – Approfondimento conoscenze specialistiche in aula.

Sub-modulo A.1 – Inglese Tecnico

Il corso mira a far sviluppare le abilità di comunicazione e di comprensione di materiale scientifico in lingua inglese. L'analisi di testi scritti a carattere scientifico inerenti le tematiche dell'IT e del project management completerà il programma formativo del sub-modulo.

Numero di ore: 25

Contenuti:

- Nozioni di grammatica
- Listening and Speaking
- Lettura e analisi di testi tecnici e articoli scientifici.

Sub-modulo A.2: Tecniche avanzate per lo sviluppo di applicazioni Java

Obiettivo principale del corso è di formare programmatori esperti alla progettazione e realizzazione di sistemi software complessi. Particolare attenzione è dedicata anche alle problematiche legate all'efficienza ed alla concorrenza di applicazioni sviluppate in linguaggio java.

Il corso fornisce un'approfondita conoscenza di tecniche e concetti utili all'analisi ed alla progettazione di software, con particolare riferimento alla programmazione ad oggetti.

Numero di ore: 75

Contenuti

- 1) Il linguaggio Java: feature avanzate
 - Exceptions, logging, assertions, and debugging
 - Generic programming, Annotations and Collections
 - Streams and Files
 - XML and StAX API
 - Internationalization
 - Security
 - Compiler API and Native methods
- 2) Strumenti e metodologie
 - Workbenck e sviluppo cooperativo
 - Sviluppo guidato dai test
 - Strumenti per la gestione ed il build del software (Maven?)
 - Tools for Testing, Debugging e Profiling
- 3) Best Practices and Design Patterns

- Progetto di Architetture;
- Patterns e anti-patterns;
- Agile Software Development;
- 4) Multithreading and Distributed Programming
 - Thread e thread-safety
 - Race conditions
 - Sincronizzazione
 - Networking and Database Access
 - RMI and Web services
 - JNDI and LDAP
- 5) Application Development Framework for Enterprise Java
 - Java beans
 - Dependency Injection and Bean Wiring
 - Aspect Oriented Programming
 - MVC Frameworks
 - Persistency

Sub-modulo A.3 – Rappresentazione e Gestione della Conoscenza

La problematica affrontata dal corso abbraccia tematiche di rappresentazione della conoscenza tramite ontologie. In particolare, saranno presentati linguaggi e strumenti per lo sviluppo di applicazioni knowledge-based.

Numero di ore: 75

Contenuti:

- Introduzione al Data-Modeling
- Basi di dati
 - Entity-Relationship model, Unified Modeling Language
 - SQL - XML-based query languages
- Ontologie e Web Semantico
 - Basi di Dati vs Ontologie
 - Logica e Ontologie: Description Logics (basics) e Rule Languages (DLP)
 - Linguaggi per Ontologie: W3C Standards (RDF(S), OWL) e Rule Based (Flogic, OntoDLV)
- Progettazione di Applicazioni Knowledge-based
 - Metodologie
 - Modeling tools: editor XML e UML ed Editor di Ontologie (Protegé, Ontobroker, Ontodlv)
 - DL systems and and rule reasoners: Pellet, Kaon, Ontobroker, Ontodlv
 - DIG interface for Description Logics systems
 - API for Logic Programming
- Applicazioni
 - Reasoning on top of Databases (deduzione, data integration)
 - Reasoning on top of Ontologies (Ontology-based Information extraction)
 - Declarative Problem-Solving

Sub-modulo A.4 – Integrazione di Risorse

Obiettivo del sub-modulo è lo studio delle principali tematiche legate all'integrazione di risorse informative eterogenee e distribuite. A tal fine saranno analizzati vari modelli ed architetture di integrazione, nonché tecniche di risoluzione di eterogeneità. Il laboratorio consisterà nello studio e nell'utilizzo di componenti software per l'integrazione di dati eterogenei.

Numero di ore: 50

Contenuti:

1. Introduzione alle problematiche di integrazione di risorse:
 - Eterogeneità a livello delle tecnologie;
 - Eterogeneità a livello di schemi;
 - Eterogeneità a livello di dati;
 - Eterogeneità di servizi informativi.
2. Modelli e architetture di integrazione:
 - Federazioni (cenni);
 - Sistemi cooperativi;
 - Data warehouse;
3. Tecniche e metodi di risoluzione di eterogeneità
 - Tecniche ETL;
 - Modelli di dati e standard;
 - Modelli di mediazione: Approccio GAV, Approccio LAV, Altri approcci;
 - Riparazione di database in presenza di vincoli;
4. Tecniche di valutazione di query:
 - Definizione formale del problema di interrogazione di un sistema di integrazione;
 - Valutazione delle interrogazioni nei modelli GAV e LAV;
 - Interrogazioni in presenza di dati inconsistenti o incompleti (consistent query answering);
5. Analisi e studio di casi.

Sub-modulo A.5 - Data Warehouse e OLAP

Obiettivo del corso è quello di introdurre problematiche avanzate di progettazione ed utilizzo di sistemi di data warehouse. Una particolare enfasi è posta sui sistemi OLAP, con riferimento ai principi, modelli e strumenti per l'analisi multidimensionale ed OLAP di grandi volumi di dati. Inoltre, il corso focalizza l'attenzione sugli ambienti e sugli strumenti Open Source per il supporto di Data Warehousing e OLAP. Parte integrante del corso è rappresentata dal disegno e sviluppo di un progetto su Data Warehousing ed OLAP per l'elaborazione e l'analisi di dati multidimensionali.

Numero di ore: 50

Contenuti:

- Introduzione ai Data Warehouse: scenario, motivazioni, definizioni e concetti di base.
- Estrazione e riconciliazione dei dati Disegno di un Data Warehouse: definizione delle strutture e delle strategie di acquisizione dei dati.
- Progettazione concettuale di un Data Warehouse: il modello DFM Progettazione logica di un Data Warehouse: star schema.
- Dati e metadati. Dati informativi e dati operazionali.
- Le architetture tecnologiche e le architetture dei dati.
- Problematiche di gestione del Data Warehouse.
- Architetture e piattaforme di elaborazione per il Data Warehouse.
- Panoramica dei prodotti disponibili sul mercato.
- On Line Analytical Processing (OLAP)

- Il modello dei dati multidimensionale.
- OLAP, MOLAP, ROLAP: Data mart e report Visualizzare i dati da differenti prospettive, a diversi livelli di dettaglio: slice e dice.
- Analisi dei dati presenti in forma aggregata per un maggior livello di dettaglio: drill-down.
- Modifica della visualizzazione dei dati: rotation.
- Previsioni su dati temporali: trend analysis.
- Query e reporting.
- L'integrazione degli OLAP server nelle soluzioni di Business Intelligence:
 - Presentazione di server OLAP
 - Tool di interrogazione, reportistica e visualizzazione

Sub-modulo A.6- Data e Text e Mining

Il corso si propone di affrontare le tematiche di scoperta di conoscenza da dati strutturati e testi. A tal fine, saranno studiati i principi, i modelli, le architetture ed i principali sistemi di data e text mining. Le esercitazioni in laboratorio saranno dedicate all'analisi ed all'utilizzo di sistemi sw specializzati per il mining di dati strutturati e testi.

Numero di ore: 75

Contenuti:

1. Data Mining

- Introduzione al data mining; Applicazioni, Ciclo di KDD, metodologie di DM. CRISP-DM.
- Data Understanding & Preparation: Weka, Rialto e R, Tipi di dati; Accesso ai Dati; Sampling.
- Data Exploration: Istogrammi e distribuzioni, Scatter Plots, Box Plots; Summary Statistics.
- Data Preparation & Cleaning: Valori anomali, valori mancanti, Outliers, Conversioni.
- Casi di studio: Churn, Algae Bloom.
- Modelli di Data Mining. Alberi di decisione, Nearest neighbor, Support Vectors Machines. Valutazione.
- Decision Theory e teoria delle probabilità; Distribuzioni di probabilità; stima: interval vs. point estimation. Maximum likelihood.
- Modelli lineari. Regressione lineare e regressione Logistica. Interpretazione probabilistica.
- Metodi Ensemble. Adaboost.
- Metodi Ensemble e Mixture Models. Clustering.
- Casi di studio (Fraud Detection in ambienti estremamente imprecisi, Propensione agli investimenti in ambiente bancario, ROC Analysis, Lift/Gain charts, Churn analysis, Insurance redlining, RFM Modeling e customer segmentation)

2. Text Mining

- Pre-processing di testi, selezione di feature
- Induzione di regole di classificazione
- Sistemi di classificazione testuale
- Aspetti avanzati: Algoritmi genetici.
- Automated Text Clustering
 - A definition of the clustering task

- Applications of text (and term) clustering
- Clustering algorithms
- Evaluating clustering algorithms
- Labelling clusters

Sub-modulo A.7 - Workflow e Processi

Obiettivo del sub-modulo è lo studio delle principali tematiche legate alla gestione informatizzata di processi ed, in particolare, agli strumenti di workflow management. Il corso si focalizzerà su due aspetti. Innanzitutto, si forniranno gli strumenti concettuali necessari per affrontare le fasi di progettazione, analisi, e modellazione di processi aziendali; in tale ambito, particolare attenzione sarà rivolta a metodologie di progettazione che prevedano UML come strumento formale per la specifica di processi. Successivamente, saranno investigati aspetti più tecnologici presentando, ad esempio, alcune architetture concrete di sistemi di workflow, i linguaggi di specifica maggiormente diffusi per applicazioni commerciali, nonché alcuni meccanismi di invocazione remota di servizi. Le lezioni teoriche saranno affiancate da esercitazioni.

Numero di ore: 50

Contenuti:

- Introduzione ai Processi di Business e Workflow
 - Concetti di base
 - Classificazione
 - WFMS: WPMC reference model
- Modellazione ed emanazione di Workflow:
 - Reti di Petri
 - Workflow Patterns
 - YAWL (Yet Another Workflow Language)
- Workflow Analysis
 - Teoria delle code per analisi
 - Strumenti basati sulla Simulazione
- Business Process Intelligence per l'analisi ed il re-engineering di processi
 - Introduzione alla rappresentazione ed all'analisi dei log di esecuzione di un processo
 - Tecniche per la scoperta di pattern frequenti
 - Algoritmi per l'induzione di modelli workflow
 - ProM
- Validazione basata sui Log dei modelli di processo: conformance checking, verifica di proprietà logiche (LTL).
- Laboratorio:
 - utilizzo di tool per il Business Process Management: YAWL, ProM e ProM-Import.

Sub-modulo A.8 Introduzione alla Service Innovation, modelli e tools

Obiettivo del sub-modulo è di presentare e chiarire gli aspetti chiave del modello di sviluppo software basato sulla Innovazione aperta tramite l'introduzione nel ciclo di vita di sviluppo del software degli utenti finali. Si analizzeranno modelli, tools e best practices, in particolare relative tecnologie abilitanti i modelli self-service e on-demand (Cloud Computing languages e tools, Virtualizzazione, etc), le caratteristiche e le limitazioni nel disegno, implementazione ed erogazione di servizi al cliente. Inoltre saranno studiati diversi scenari implementativi e come il modello Cloud soddisfa requisiti espliciti ed impliciti nei vari settori di mercato (PA, Media, Sanità, Industria, ecc).

Numero di ore: 40

Contenuti:

- Introduzione alla Service Innovation
 - Open Innovation e User-Centered Sw Development
 - Modelli e tools di sviluppo software basato sulla Open Innovation
 - Progettare software ed architetture innovative
- Introduzione al Cloud Computing
 - Terminologia e modelli (IaaS, PaaS, SaaS)
 - Pro e Contro del modello a servizi
 - Opportunità e sfide – situational computing
- Analisi dei modelli e processi sottostanti alle offerte commerciali e soluzioni Open Source
 - Software as a Service (SaaS) – Google Apps, SalesForce
 - Platform as a Service (PaaS) – Google App Engine e Microsoft Azure, Force.Com, Heroku, CloudFoundry, CloudForce
 - Infrastructure as a Service (IaaS) - Amazon Elastic Compute Cloud, RHEV, OpenStack, OpenNebula, Eucalyptus
- Modelli di deploy di soluzioni cloud
 - Cloud pubblici, privati e ibridi
 - Analisi delle architetture di riferimento
- Requisiti del Cloud Computing per vari settori di mercato
 - Privacy - data protection e data liberation
 - Security
 - Sustainability – green computing
 - Legal Compliance – le legislazioni EU

Sub-modulo A.9 Progettazione e Sviluppo di Servizi Innovativi

Obiettivo del sub-modulo è di fornire gli strumenti teorici e pratici per iniziare lo sviluppo secondo il modello dell'Open Innovation di servizi software innovativi. Il modulo conterrà miniprogetti ed esercizi volti a sperimentare le varie fasi di sviluppo di tipo aperto.

Numero di ore: 60

Contenuti:

- Strutture e modelli di piattaforme di sviluppo collaborativo
 - Piattaforme sociali e living Labs
 - Testbeds e eScience 2.0
 - Software life-cycle Management collaborativo
- Framework Commerciali per la realizzazione di soluzioni cloud
 - Protocolli di comunicazione (SOAP, REST) e integrazione con applicazioni esistenti
 - L'offerta di Microsoft Azure (Windows Azure Compute: Web Role and Worker Role; Windows Azure Storage: Blob, Drive, Table, Queue, SQL Azure; App Fabric: Caching, Access Control, Service Bus)
 - L'offerta di Google App Engine (Google Apps, SDK e Eclipse Plugin, Data Store, URL Fetch, Task Queues, Mail, Channel, Google Web Toolkit)
 - L'offerta di Amazon (S3 Simple Storage Server, Simple Queue Service).
- Il futuro del Cloud Computing
 - standardizzazione e iniziative per un cloud aperto

- Bandi EU e prospettive di Ricerca
- Future Internet

Sub-modulo A.10: Elementi di Bioinformatica e gestione ed analisi di dati omici

Gli obiettivi del modulo sono di fornire le competenze di base sulle metodologie e gli algoritmi della bioinformatica e le principali tecniche di rappresentazione ed analisi data mining di dati estratti da piattaforme biotecnologiche (dati omici). Al termine del corso gli studenti conosceranno i principali dati omici e saranno in grado di eseguire un esperimento completo di classificazione di dati omici utilizzando piattaforme open source di data mining accoppiate a strumenti per il preprocessing di dati omici.

Numero di ore: 50

Contenuti:

- Elementi di Bioinformatica
- Rappresentazione di biomolecole (DNA, proteine)
- Banche dati biologiche (GenBank, UniProt, PDB, ecc.)
- Ricerca in Banche dati biologiche
- Allineamento di sequenze locale e globale
- Multiallineamento
- Rappresentazione secondaria e terziaria di proteine
- Predizione e allineamento di strutture
- Introduzione alla genomica e alla proteomica computazionale
- Rappresentazione informatica di dati genomici e proteomici
- Preprocessing e analisi data mining di dati microarray (gene expression, genotyping) e mass spectrometry
- Rappresentazione ed analisi di reti biologiche (es. interattomica)
- Utilizzo di ontologie nell'analisi dei dati omici
- Introduzione a Gene Ontology
- Casi studio in medicina e biologia: biomarker discovery e applicazione della classificazione a dati genomici e proteomici

Sub-modulo A.11: Elementi di Informatica Medica e gestione ed analisi di dati clinici

Gli obiettivi del modulo sono di fornire le competenze di base sulle metodologie e gli algoritmi dell'informatica medica, i principali standard e formati per la rappresentazione e trasmissione di dati clinici e le metodologie e strumenti per l'estrazione di conoscenza da dati medici e sanitari. Al termine del corso gli studenti conosceranno i principali standard medici e saranno in grado di manipolare dati clinici e sanitari e saranno in grado di progettare delle semplici cartelle cliniche elettroniche o sistemi informativi sanitari di supporto alla gestione di biosegnali e bioimmagini.

Numero di ore: 50

Contenuti:

- Elementi di informatica medica,
- Principali standard per la rappresentazione di dati in medicina e sanita (es. ICD, SNOMED, LOINC)
- Principali standard per la trasmissione su rete di dati in medicina e sanita (es. DICOM, HL7)
- Introduzione alle bioimmagini (es. TC, MRI, PET)

- Sistemi per la gestione di bioimmagini (RIS, PACS)
- Introduzione ai biosegnali (es. ECG, EEG)
- Sistemi per la gestione di biosegnali
- Metodologie per la rappresentazione, gestione ed analisi data mining di dati medico-clinici
- Cartelle cliniche elettroniche
- Sistemi informativi ospedalieri
- I principali processi nei sistemi informativi ospedalieri (es. SDO, DRG)
- Casi studio: Scoperta di conoscenza da Cartelle cliniche elettroniche e Sistemi informativi ospedalieri

Sub-modulo A.12: Sistemi ad alte prestazioni per l'analisi data mining distribuita

Il corso intende fornire le metodologie, le tecniche e gli strumenti inerenti il calcolo parallelo ed i sistemi distribuiti, con particolare riferimento ai sistemi Grid e Cloud, illustrando i benefici della loro applicazione a problemi nel settore del data mining.

Al termine del corso gli studenti conosceranno le principali architetture parallele ed i principali modelli di calcolo parallelo e distribuito e sapranno utilizzare piattaforme di data mining distribuiti e a servizi.

Numero di ore: 50

Contenuti:

- Introduzione alle architetture parallele
- Architetture Multicore, Supercomputer, Multicomputer, Cluster
- Introduzione al Calcolo parallelo
- Modelli di Programmazione parallela
- Cenni agli Ambienti PVM, MPI,
- Programmazione parallela in Java
- Service Oriented Architecture
- Introduzione al Grid Computing per elaborazione massiva di dati
- Principali Ambienti per il Grid Computing (Globus)
- Introduzione al Cloud Computing e ambienti per la programmazione di applicazioni data-intensive su rete
- Modello MapReduce, Ambiente Apache Hadoop
- Esempi di applicazioni scientifiche parallele
- Introduzione al Data Mining Parallelo e Distribuito
- Principali piattaforme software per il data mining parallelo, distribuito e a servizi
- Esempi di analisi data mining distribuita

Sub-modulo A.13: Lavoro di tesi

Numero di ore: 50

Contenuti:

Obiettivo del sub-modulo sarà la redazione di un lavoro di tesi dedicato all'approfondimento di uno dei temi trattati durante il Modulo A. I partecipanti del progetto di formazione saranno suddivisi in piccoli gruppi ognuno dei quali sarà seguito da un docente-tutor. Ogni studente predisporrà una tesina che sarà oggetto della valutazione finale.

1.3 Modulo B – Apprendimento di conoscenze di programmazione, gestione strategica, valutazione ed organizzazione operativa dei progetti di ricerca.

Il Modulo B prevede l'erogazione di alcuni sottomoduli finalizzati a trasferire ai partecipanti competenze specifiche in merito alla pianificazione e gestione dei progetti di ricerca e ai processi di valorizzazione dei risultati della ricerca.

Sub-modulo B.1–Pianificazione, Gestione e valorizzazione dei risultati della Ricerca e del Trasferimento Tecnologico

Il sub-modulo ha l'obiettivo di illustrare ai partecipanti conoscenze per la gestione dell'innovazione e del trasferimento tecnologico, sia per ciò che concerne gli aspetti giuridici, sia per quanto concerni gli aspetti relativi al technology management. Il Sub-modulo B1 si propone, altresì, di presentare gli strumenti finanziari e di policy a supporto della ricerca, dell'innovazione, e del Trasferimento Tecnologico.

Numero di ore: 40

Contenuti:

- La tutela giuridica del software;
- Classificazione Giuridica del Software;
- La tutela della privacy
- Il sistema della Ricerca Nazionale;
- Il sistema della Ricerca nei paesi avanzati e nelle economie emergenti;
- Il Sistema industriale nazionale in relazione alla R&D;
- Il quadro Strategico Nazionale e il PON Ricerca e Competitività;
- Gestione del Trasferimento Tecnologico.

Sub-modulo B.2– Project Management

Il sub-modulo Project Management ha l'obiettivo di trasferire ai partecipanti metodologie e strumenti pratici per gestire le attività di realizzazione di un progetto di ricerca industriale e di sviluppare metodi per condividere, comunicare e documentare le attività di realizzazione del progetto.

Risultati attesi sono l'apprendimento di una metodologia e di un linguaggio comune per la gestione del progetto, del Team, della comunicazione, dei tempi e dei costi. L'obiettivo è quello di trasferire gli schemi generali e le modalità operative che permettono di gestire i progetti in modo interattivo, intuitivo ed efficace. L'apprendimento sarà ottenuto mediante la pianificazione di un caso aziendale concreto di un progetto sviluppando la "Work Breakdown Structure" e il "Piano di Progetto" e le "modalità di avanzamento" in modo da apprendere le tecniche di Project Management anche mediante il loro utilizzo pratico.

Numero di ore: 30

Contenuti:

- La metodologia del Project Management:
 - principi del Project Management;
 - progetto;
 - processi di Gestione del progetto;
 - processo di pianificazione;
 - piano di progetto.
- Strumenti:

- Work Breakdown Structure;
- L'avanzamento e la comunicazione:
 - Sistema di reporting del progetto;
 - Avanzamento del progetto;
 - Comunicazione di progetto;
 - Chiusura del progetto;
 - Gestione del multiprogetto.
- Software di gestione.

Sub-modulo B.3– Economia e Gestione delle Imprese

Il sub-modulo di Economia e Gestione delle Imprese ha come obiettivo quello di fornire elementi di teorie economiche, modelli manageriali e tecniche di gestione, con l'obiettivo di trasferire agli allievi gli strumenti necessari a governare i processi di funzionamento dell'impresa.

Numero di ore: 30

Contenuti:

- Le teorie sulle finalità dell'impresa fra creazione, diffusione e distribuzione del valore economico;
- L'analisi delle risorse che determinano la capacità di creazione di valore dell'impresa: relazioni e conoscenza nel processo evolutivo dell'impresa quale sistema cognitivo;
- L'impiego delle risorse aziendali nella prospettiva delle finalità e degli obiettivi imprenditoriali: visione, missione, dinamiche competitive e strategie d'impresa, sui mercati nazionali e internazionali;
- Le gestione dell'innovazione e delle tecnologie per lo sviluppo aziendale - la funzione Ricerca & Sviluppo (R&D);
- Lo sviluppo imprenditoriale e il business planning: principi e processi;
- Fondamenti di marketing e di gestione commerciale;
- Principi di finanza e di gestione finanziaria dell'impresa;
- Tecniche di analisi per la gestione d'impresa.

1.4 Modulo C – Esperienze operative in affiancamento a personale impegnato in attività di ricerca (training on job).

Il Modulo C prevede una fase di training on the job presso le aziende partner del progetto di ricerca ad esso associato. Durante tale periodo gli allievi saranno chiamati a mettere in pratica quanto appreso durante le attività d'aula ed a collaborare nelle attività di ricerca e sviluppo sperimentale previste dal progetto.

Le attività di training on the job sono finalizzate a fare acquisire ai partecipanti al corso la dimensione pratico/funzionale del ruolo professionale che saranno chiamati ad assumere nel mercato del lavoro una volta formati.

Gli obiettivi dell'attività di training on the job sono:

- integrare le competenze formative acquisite e realizzate in aula con le propensioni professionali dell'utenza;
- costituire un sostegno reale e complementare alla formazione delineando in concreto obiettivi di orientamento, di educazione al lavoro, di approfondimento ed acquisizione di nuove competenze e conoscenze.