POLITECNICO DI MILANO

PhD in INGEGNERIA DELL'INFORMAZIONE / INFORMATION TECHNOLOGY - 39th cycle

Research Area n. 1 - Computer Science and Engineering

THEMATIC Research Field: MEDICAL DATA EXCELLENCE: STORING, INTEGRATING, AND PROCESSING WITH QUALITY IN MIND

Monthly net income of PhDscholarship (max 36 months)
€ 1400.0
In case of a change of the welfare rates during the three-year period, the amount could be modified.

Context of the research activity	
Motivation and objectives of the research in this field	The exponential growth of data in the medical domain highlights the necessity for structured and unified access to numerous and heterogeneous datasets. Efficiently storing and managing these datasets in a way to accommodate all the formats is vital for further analysis and applications. This project aims to address these challenges by: i) developing a unifying, comprehensive, and global catalog that encompasses various data formats and associated metadata. The unified schema repository will serve as a reference ontology, and it has to take into consideration both technical and clinical information. ii) empowering users with the capability to seamlessly navigate, search, and explore all available datasets, along with the related ontology and metadata. The research will support users in finding the datasets that, alone or integrated, would provide the greatest value considering the designated goals. iii) supporting the design of an effective data preparation pipeline for guaranteeing high-quality data as input to the data analysis step. The goal is to find the most appropriate sequence of data preparation techniques in order to maximize the performance of the data analysis algorithm and minimize the preprocessing time.

Methods and techniques that will be developed and used to carry out the research	The student will focus on different aspects of data management: data architecture, data quality and data integration. The objective is to design a medical data repository in which datasets have to be described in terms of metadata that have to model both technical and clinical information. The students will have to focus the research on data quality techniques and semantic data integration.
Educational objectives	The student will learn how to store and manage heterogeneous datasets. The student will also learn how to prepare datasets in order to guarantee high quality data for the users analyses.
Job opportunities	Data management, governance and especially data quality are attractive research fields and professional areas. Graduates with expertise in this area can easily find jobs both in academia and in private organizations.
Composition of the research group	1 Full Professors 2 Associated Professors 2 Assistant Professors 2 PhD Students
Name of the research directors	prof.ssa Cinzia Cappiello

Contacts
cinzia.cappiello@polimi.it

Additional support - Financial aid per PhD student per year (gross amount)	
Housing - Foreign Students	--
Housing - Out-of-town residents (more than 80Km out of Milano)	--

Scholarship Increase for a period abroad	
Amount monthly	700.0 €
By number of months	6

Additional information: educational activity, teaching assistantship, computer availability, desk availability, any other information
EDUCATIONAL ACTIVITIES (purchase of study books and material, including computers, funding for participation in courses, summer schools, workshops and conferences): financial aid per PhD student.

TEACHING ASSISTANTSHIP: (availability of funding in recognition of supporting teaching activities by the PhD student) There are various forms of financial aid for activities of support to the teaching practice. The PhD student is encouraged to take part in these activities, within the limits allowed by the regulations.

COMPUTER AVAILABILITY: individual use

DESK AVAILABILITY: individual use