

TESI @ UNIVERSITAT POLITÈCNICA DE CATALUNYA – BARCELONA (SPAIN)



The **Research and Development Lab (Rdlab)** manages directly all the software and hardware resources of the Computer Science department research groups. The available computing infrastructure provides:

Over 160 physical servers.

Over 1000 CPU cores and more than 3 TBytes of RAM memory.

Over 130 TBytes of disk space.

High speed network at 10Gbit..

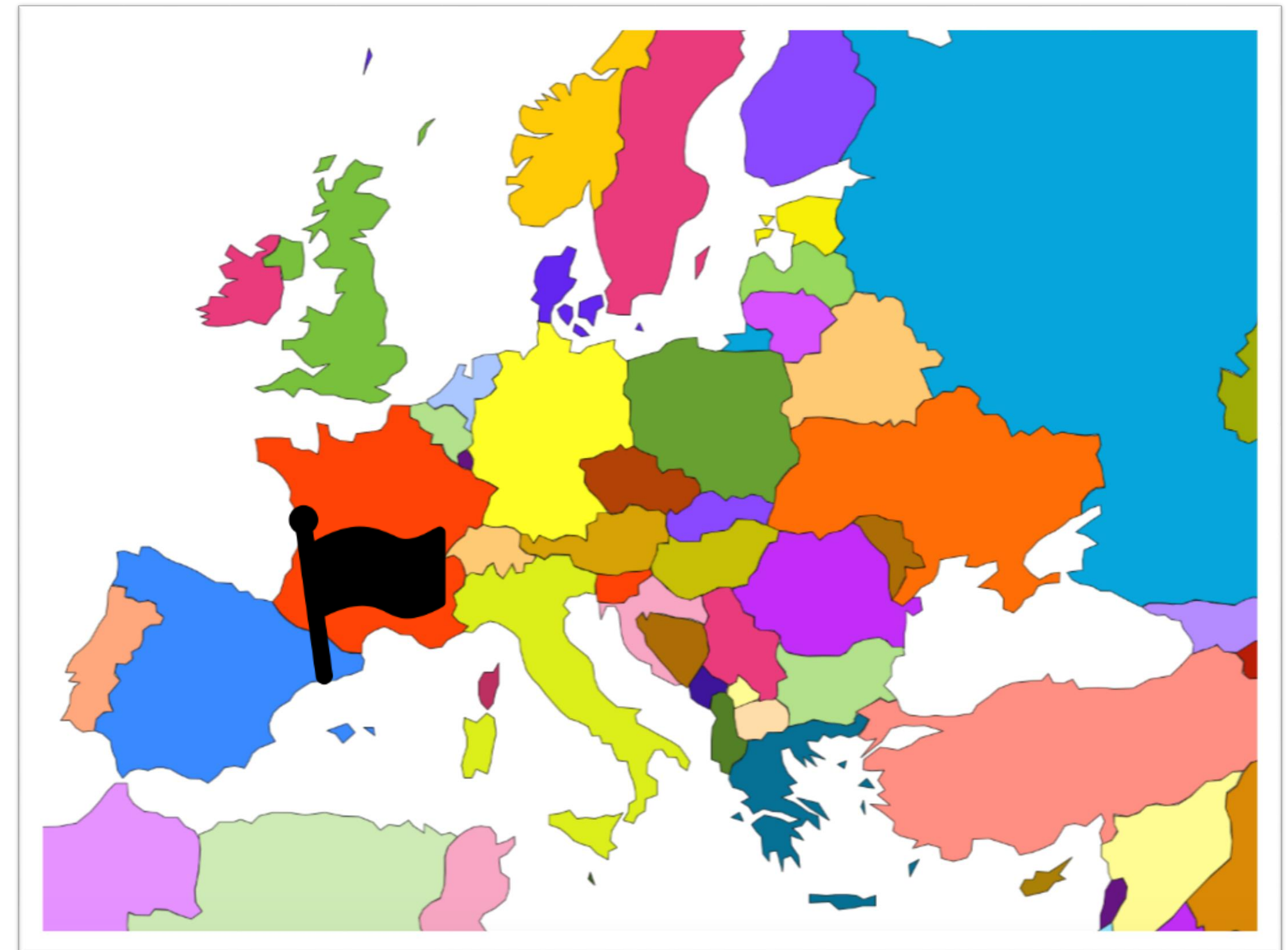
Fatos Xhaifa, the contact professor for the ERASMUS@UPC-BarcelonaTECH, has been awarded for research merits for the six year periods 1996-2001-2007-2013 (tres sexenios) by CNEAI, Spanish Ministry and for the periods 2001-2006-2011 (two sexenis compl. autonòmic.). His research/teaching activities in the fields of Distributed and Parallel programming, Parallel Architecture, P2P and Grid Computing is evaluated to A/A by the Universitat Politècnica de Catalunya.

Prof. Fatos in numbers: H-Index: 45; Citations: 8996

Required Skills for the thesis:

Machine Learning e Big Data Analytics

Give a look to the Lab Website



Andrea Giordano



Universitat Politècnica de Catalunya

GRAPH MATCHING PER LA RICERCA DI PATTERN IN DATABASE BIOLOGICI: STIMA DELLE PERFORMANCE ED OTTIMIZZAZIONI

Abstract



Today, one of the most interesting research fields in IT industries is the quest for software able to deal efficiently with Big Data and taking advantages coming from the use of applications which produce them. Some of them are so performant that are largely used from the most important Big Data dealing companies in the world like Netflix, Facebook and Google.

A side aspect but not less importantly is related to the infrastructure generally used in conjunction with these technologies, i.e. distributed computing, which exploits the power of several, usually a lot of, physical machines to perform very complex tasks with high performances outcome. In this thesis firstly we are going to study Big Data panorama as a whole, especially about their spread in the IT industries and in Healthcare, where data are generated at a high rate and the use of computer science can save or improve lives. Moreover, the challenges and a vision of the potential uses of them are rapidly covered together with a view on so-called wearables: sensors applied on clothes or directly on people's skin which already are revolutionizing our lives. In this dissertation the Introduction and the chapter 2 are dedicated to these persons.

The aim of this project is to design and develop a system composed of a cluster intended for medical use where the core is a real-time detection of anomalous pattern within data originated by wearables. A lot of papers and technical reports of scientific community were consulted to investigate the State of the Art of Big Data handling technologies: studying their abilities and properties, finding strengths and weakness of each one and comparing them in order to choose the most adequate set of software to achieve the aim of the project.

Therefore, an original architecture is designed assembling several technologies and taking into account requirements and targets to reach: every design choice is justified in the text and a particular attention was paid to assure the essential IT properties of modularity and simplicity without renouncing to high performances. The sections 5.3 - 5.6 present the involved devices and some contributes brought to already existing tools; additionally, an original approach to handle semantic-enriched data streams is illustrated, taking inspiration from an innovative way submitted at the latest ISWC conference. Furthermore, the document presents the development of a prototype based on the designed system and shows the outcomes of some experiments performed on it: these ones concern resource's usages in different operating conditions on the cluster nodes. In chapter 5 are showed also the results obtained using the employed anomaly detection algorithm. Finally, Future works proposes potential improvements adding a cluster to execute semantic analysis in order to reach higher achievements about medical diagnosis and data comprehension.

The chapter 8 lists the most celeb existent technologies to do it and shows how their integration with the presented system can provide substantial advantages in Healthcare as well as illustrates obstacles and challenges to overcome by presenting a good starting point to implement further developments.