MAP-i 2011-12

Pre-thesis & Free Option Planning

PhD Student:	**Sara dos Santos Fernandes**
Theme:	**Exploiting the FLOSS paradigm in Software Engineering education**: **Design, monitoring and validation of a collaborative e-learning platform**
Supervisor:	*Antonio Cerone*, `www.iist.unu.edu/~antonio` International Institute for Software Technology, United Nations University *Luis Soares Barbosa*, `www.di.uminho.pt/~lsb` INESC TEC and Universidade do Minho
MAP-i Center:	HASLab - High Assurance Software Laboratory, INESC TEC `haslab.di.uminho.pt`

1 Research context & objectives

1.1 The problem

Modern IT-driven societies demand highly skilled professionals who can successfully design complex systems at ever-increasing levels of reliability and security. Both the global, world-wide nature of this demand and the need to empowering developing countries to meet the challenges of nowadays and future information society at all levels (administrative, social and economic) entail the need for the development of innovative, light, cheap, adaptable but effective frameworks for high-level training in Software Engineering.

Contrary to traditional education practices, such frameworks would stress learner's responsibility and control over his/her own learning process, encourage the development of cross-cutting competences rather than the mere acquisition of content knowledge, and be essentially project-driven in highly interactive environments. These concerns are in line with current research on Education which recognises that learning and development are fundamentally the results of participation in social interactions and culturally organised activities with others. It is also recognised the underlying potential of digital information and communication technologies to mediate learning experiences and support networked collaborative learning.

1.2 Aims

Commons-based peer-production is the model of economic production in which the creative energy of large numbers of individuals is remotely coordinated, usually through the Internet, into large, meaningful projects mostly without traditional hierarchical organization. Typical instantiations of peer-production models are Free/Libre Open Source Software (FLOSS) projects.

The acronym FLOSS refers to any software distributed under a license allowing users to change or share the software source code. Along a complex history, marked by the increased availability of computational resources and debates over the nature of knowledge and information, FLOSS changed the way societies think about software and have come to inform a cultural phenomenon that is underpinned by technological development with the aim of contributing to the public good.

This PhD project aims at exploiting FLOSS based approaches to build, monitor and validate networked learning environments for high-level training in Software Engineering. Its working hypothesis is that distributed collaborative networks working together in a demanding software project can provide a powerful learning environment to complement (or even to act as an effective alternative to) formal education practices. The project will develop an e-learning framework for teaching Software Engineering soft and hard skills within a peer-production environment that fosters collaborative learning and allows students to enable effective learning of Software Engineering through strong participation in real-world software projects.

The e-learning framework will be applied on a number of geographically distributed pilot studies based on existing FLOSS projects in which students participate both as learners, by exploiting the FLOSS community as an e-learning environment, and as researchers, by analysing the FLOSS phenomenon in terms of its learning capabilities.

It will be developed a specific pilot study based on the creation of a new FLOSS project in which the FLOSS paradigm is extended by incorporating innovative technologies, usability support and e-Learning tools to facilitate students activities, and their use as pilot projects in formal curricula. Such innovative technologies typically include but are not restricted to formal software development methods, source code analysis and reverse engineering techniques.

The PhD project will encompass the design of this framework as well as the monitoring, analysis and assessment of the pilot case studies to validate the working hypothesis and elaborating methodological and technical guidelines to pursue this line of research.

1.3   Context

This PhD project is funded by the International Institute for Software Technology of the United Nations University (UNU-IIST) which is located in Macau, in the context of the PPAeL project on *Peer-production approaches to e-Learning* led by Dr Antonio Cerone.

The mission of the UNU International Institute for Software Technology is to foster the development and application of information and communication technology to address the pressing global challenges of sustainable development that are the concern of the United Nations, its Peoples and Member States through education, research, capacity development, and policy support. During the 20 years of its history, UNU- IIST has addressed this mission by carrying out research activities in various disciplines that include software engineering, electronic governance and education, by delivering education to developing countries through courses, schools, workshops and supervision of postgraduate students and by building an international network of collaborations with academic institutions and government and non-government organisations all over the world. It aims to be an intellectual focal point that attracts the top, most innovative minds in research and teaching, and trains students to become leaders of change.

Pilot studies will be carried on at Universidade do Minho (Portugal), Thessalonik University (Greece) and at least a developing country outside Europe.

2   External member of the thesis committee

One external member of the thesis committee will probably be selected among the research staff of the *Istituto per le Tecnologie Didattiche - Genova, Italy,*

3 Pre-thesis Planning

3.1 Aims

1. To assess and rephrase the problem statement objectives originally stated in this proposal, clearly identifying the relevant research questions;

2. To support the proposed research with preliminary field studies and state of the art review;

3. To propose a detailed thesis plan and schedule of research activities.

3.2 Activities and outcomes

State of the art review. This is to be done in two main areas:

- Collaborative learning and e-learning support frameworks (methodologies, tools, experiences)

- FLOSS projects used in formal or informal educational contexts with a particular emphasis on possible/existent relationships with Software Engineering curricula development.

Deliverable: part of the pre-thesis report (with suitable and extensive bibliographic references)

Preliminary research on FLOSS projects as learning environments. Conduct a preliminary enquiry to a selected group of FLOSS performers on their own experience to identify emerging learning patterns along FLOSS development. This activity will be prepared with the usual tools of research on Education (e.g. a questionnaire followed by interview to selected cases) and applied to restricted number of people, preferably with different profiles.
Deliverable: part of the pre-thesis report, eventually leading to a publication.

Preliminary framework conceptualisation and design. The experience gathered in the previous activities are supposed to lead to the development of a prototype of the e-learning framework to be tested and validated in a preliminary pilot case study, including some innovative technologies, such as formal methods, to launch in October 2012. This will act as testbed for framework design and implementation to be accomplished along the first non curricular year of the PhD programme.

Deliverable:

- Prototype of the e-learning framework

- Design of a preliminary pilot study (restricted to a single course unit)

4 Free Option Planning

4.1 Aims

1. To become proficient in some areas/topics, understood as a necessary background for the proposed plan of doctoral studies.

4.2 Description

Data analysis and ontologies. Communication and patterns of contribution in FLOSS projects have emerged as an important measure in assessing the quality of FLOSS products. More recently, it has been noticed that such factors have also a strong impact on the learning process that naturally occurs through individual participation in a FLOSS project as well as on the global knowledge accumulated within the corresponding FLOSS community. Collaboration in FLOSS projects is highly mediated by the usage of tools, such as versioning systems, mailing lists, reporting systems, etc. These tools serve as repositories which can be data mined to understand the identities of the individuals involved in a communication, the topics of their communication, the amount of information exchanged in each direction, as well as the amount of contribution in terms of code commits, bug fixing, reports and documentation produced and email postings. These large amount of data can be selectively collected and then analysed not only by using inferential statistics to identify activity patterns but also by using ontology engineering formalisms that support the extraction of semantic information.

The aim of this activity is to review existent work on ontologies and social network analysis concerned or relevant to the systematic classification, visualisation and analysis of the above mentioned data (which is supposed to be collected through suitable data mining procedures in another component of the PPAeL project.).

Deliverable: research notes; proposal of a suitable ontology for classification of FLOSS emerging learning patterns and related data.