Erasmus Mundus Mobility with Asia

ERASMUS MUNDUS ACTION 2

N° of grant agreement: 2013-2539 / 001-001- EM
Action 2 Partnerships

EMMA: ERASMUS MUNDUS MOBILITY WITH ASIA

DOCTORAL RESEARCH PROJECT

RESEARCH PERIOD:

Name of the student: Jane Doe	**Emma Number 20140033**
Country: LAO PDR	**Email :** jane.doe@gmail.com

Doctoral School: ED STIC	**Host laboratory:**
Name of the advisor:	**Email of the advisor**

Title of the thesis (provisional): Biodiversity Knowledge extraction Techniques (BioKET: Data Mining) **Field of research:** Computer Science

Description:

This work addresses several major problems of *biodiversity data mining*: How can we integrate heterogeneous data and knowledge in the data mining process? How can we make the best use of the available biological knowledge for optimizing the relevance of extracted knowledge patterns and models? How can we represent, compare and integrate the extracted patterns in ontologies and biological knowledge bases?

The first part of this work consisted into two tasks: First, constitute at the same time a state of the art of health, social and economic impacts, and applications of data mining in biodiversity, and an annotated bibliographic repository of publications and data/knowledge resources in biodiversity data mining. Second, create the database integrating botanical and medical data on plants found in South-Est Asia originating from several data sources (BRAHMS, NAPIS, BioTIK, ITM-FOF, etc.). For the realization of these steps, Mr. INTHASONE performed a presentation of our initial work and of the data mining approaches developped in the KEIA team, and conducted discussions and exchanges with several research groups in Laos.

The second part of this work will consist in the adaptation and development of semantic data mining techniques for the analysis of biodiversity data. We will be particularly interested in Galois lattice based solutions such as hierarchical *conceptual clustering*, the extraction of *conceptual association rules* and the generation of *conceptual classifiers*, i.e. the extraction of multilevel association and classification rules and patterns using generalization and specialization operations on hierarchical biological data structures, and the integration of biological knowledge - from bio-ontologies, bio-knowledge bases, bibliographic repositories, etc. - in the different stages of the data mining process in order to optimize the relevance of extracted knowledge patterns and models. Extracted knowledge patterns and models can then be used to enrich the biodiversity knowledge bases and databases.

Further useful remarks: Keywords: Data mining; Knowledge extraction; Phenotypes identification; Biodiversity studies.

Student's signature: **Date**:

Advisor's signature: **Date:**