|  |  |  |
| --- | --- | --- |
|  | **Trabajo Fin de Máster** | **Clave** |
| **TFM-A** | *Planificación y presupuesto* | **UO241069** |

***Datos Personales del Alumno/a***

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Apellidos** | **LAM** | | | | | | | **Nombre** | | **NGOC TRAN** | | | | | | | | |
| **DNI** | **Y3051458Q** | **Año de inicio de los estudios** | | | | | | | **2013** | | **Edad** | | | | | | **25** | |
| ***Domicilio*** | | | | | | | | | | | | | | | | | | |
| **Calle** | **Gonzalez Besada** | | | **Nª** | | **36** | **Piso** | | **03** | **Letra** | | |  | | **Esc.** | | |  |
| **Población** | **Oviedo** | | | **Municipio** | | | |  | | | | | | | | **C.P.** | |  |
| **Provincia** | **Asturias** | | **Teléfonos** | | **1** | **692336696** | | | | | | **2** | |  | | | | |
| **Email** | **UO241069@uniovi.es** | |  | |  |  | | | | | |  | |  | | | | |

**DATOS DEL TRABAJO FIN DE MÁSTER**

|  |
| --- |
| **Título: CROWDSEM - A crowdsourcing based system for multilingual linked open data** |
| **Autor:Lam Ngoc Tran** |
| **Director/a: Professor Jose Emilio Labra Gayo** |
| **Rama Profesional** |

|  |
| --- |
| **Project Purpose or Justification:** |
| The purpose of the project is to develop a web system (**CROWDSEM**) which can automatically:   1. Retrieve, process, validate and correct semantic web data. 2. Translate semantic web data in several languages.   Besides, the system needs to provide easy-to-use interfaces to attract and collect answers from volunteers. These interfaces should work in different devices such as desktop, laptop, tablet and mobiles. |

| **Project Description:** |
| --- |
| Semantic web is one of the popular trends of web development. The idea is to convert current web into a “web of data” which allow data to be shared and reused across application. Many semantic web services provide huge amount of information ( DBpedia, Spanish DBpedia, data.gov.uk). However, these data can be incorrect, misuse or even contradict to each other. Furthermore, information is described in sematic web using predicates, for example:  <http://biology.uniovi.es> **dc:creator** <http://uniovi.es/people#Juan>.  These predicates mostly are in English. Thus, people who not use English as their main languages can have problems to read the data. These two problems cannot be solved by machines without the knowledge of humans. Therefore, **CROWDSEM** is created to overcome these problems. The purpose is to create an environment where people can take part in a voluntary way to contribute their knowledge to correct data and to translate data from one language to other languages. By asking the same question to different people, **CROWDSEM** will automatically collect the answers, aggregate them to produce the correct result. That is the idea of Crowdsourcing  The **CROWDSEM** system will be developed in six months, from January to June. The defense of the project will be in July, according to the calendar of the Master in Web Engineering course.  The team involves 2 people: one student and one professor as a supervisor. |

|  |
| --- |
| **Project and Product Requirements:** |
| Project requirement:   * Knowledge in Semantic Web and Crowdsourcing system. * Knowledge in Web services, Programing, User interfaces, Database, Data transformation. * Two Linux servers: production server and test server.   Product requirement:   * The availability of **CROWDSEM** system in 24/7. * The data must be update frequently and up-to-date. * User can export and download the data. * Easy-to-use interfaces for different devices. * The ability to create and suggest questions to users. * The ability to collect, aggregate answers to validate and correct data. * The ability to display data in different languages. |

| **Acceptance Criteria:** |
| --- |
| * The CROWDSEM system able to query data from popular providers like DBpedia. * User can download data from CROWDSEM system in different format: JSON, CSV, and XML. * The system must have an ability to automatically verify and correct 80% of incorrect data. * The system must provide data at least in both Spanish and English. * The duration of developing must no excess six months. |
| **Initial Risks:** |
| * Availability of input data: Some SPARQL endpoints might not available at the requested time. * Availability of servers: Server can be shutdown, stop working. * Answers from spammers could make the system verify data incorrectly. * There are not many volunteers who help to verify data. Thus, the system cannot aggregate answers and verify data efficiently. * Cannot connect to other third parties services to support in translating for new data. |

|  |  |
| --- | --- |
| **Project Objectives** | **Success Criteria** |
| **Scope:** |  |
| 1. CROWDSEM service can query data from popular SPARQL endpoints.  2. The system can correct data automatically.  3. The system can display in many languages. | 1. The CROWDSEM system able to query data from DBpedia  2. The system can verify and correct 80% of incorrect data.  3. The system can display data at least in both Spanish and English. |

|  |  |
| --- | --- |
| **Time:** |  |
| The duration of a project is in 6 months | The application is deployed before 6 months. |

|  |  |
| --- | --- |
| **Quality:** |  |
| Information should be up-to-date and correct.  System functional properly | Information should be not incorrect, inconsistent or contradict.  Testing and evaluating should be conducted after each phase of the project.  Time for evaluation and testing must greater than 50% of the total duration. |

|  |  |
| --- | --- |
| **Summary Milestones** | **Due Date** |
| Research and training | 30-01-2015 |
| Collect requirements and make a final review with the supervisor before starting the project. | 15-02-2015 |
| Analysis and Design system | 22-02-2015 |
| Analysis and Develop Database | 28-02-2015 |
| Develop the CROWDSEM system to display data in many languages. | 30-03-2015 |
| Test and evaluate CROWDSEM system | 05-04-2015 |
| Develop the CROWDSEM system to validate and correct data. | 30-04-2015 |
| Test and evaluate CROWDSEM system. | 05-05-2015 |
| Integrate all modules to the CROWDSEM system. | 30-05-2015 |
| Test and evaluate the whole system. | 10-06-2015 |
| Final testing and review, documentation. | 30-06-2015 |
| Presentation | 01-07-2015 |

|  |
| --- |
| **Estimated Budget:** |
| 1. **Hardwares:** 2. 1 linux server: 30 euros for 6 months. 3. 1 mobile: 200 euros 4. 1 tablet: 250 euros 5. 1 laptop: 600 euros 6. **Salary:** 7. 1 developer (student) : 600 euros/month – 3600 euros in 6 months 8. 1 advisor (professor): 800 euros/month – 4800 euros in 6 months. 9. 30 volunteers people to test and answer questions: 10 euros/person – 300 euros in total. 10. **Other expenses:** 11. Electric, water: 50 euros / month – 300 euros in 6 months 12. Mobile bill : 200 euros in 6 months 13. Transportation: 400 euros. 14. 10% of total cost for risk management   **In total:**  **10,680 euros (without risk) – 11,748 euros (10% risk management).** |

En Oviedo, a \_28\_\_\_\_ de \_\_\_\_\_\_Enero\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ de 2015\_\_\_

|  |
| --- |
| **El alumno/a** |
| **Firma**  **Lam Ngoc Tran** |