

Perfil del Trabajo de Diploma

Curso docente 2015 - 16

Título: *Backend en Qt y python para el diccionario léxico Wordnet en Español*

Autor: *<nombre>*

Tutor: *Ing. Abel Meneses Abad*

Clasificación: *Ciclo completo.*

Clasificación del área de desarrollo: *Aplicaciones para el área del procesamiento del lenguaje natural español.*

Síntesis del Tutor: *Ingeniero en Telecomunicaciones y Electrónica (CUJAE, 2004). Profesor Asistente (UCI, 2009). Ha recibido los diplomados de Docencia Universitaria (2007), Software libre (2008), Realidad Virtual (2010), Ciencias de la Computación (2012), y Señales y Sistemas (2015). Es aspirante a doctor desde el 2014, en temas de Procesamiento de Lenguaje Natural. Miembro del Grupo de Investigaciones de Migración a Software Libre de la UCI (2004 – 09). Conferencista de Software Libre, y ha atendido la capacitación y la migración en diferentes centros de estudio e investigación de la República de Cuba. Investiga la historia del software libre en Cuba. Miembro de la Asociación Cubana de Usuarios de Tecnologías Libres. Director de Centro de Desarrollo(2010 – 13) en la Universidad de las Ciencias Informáticas. Perteneció a diversos organismos gestores de la informatización: Grupo Técnico y Grupo de Capacitación Nacional de SWL (2004 – 09). Investiga en temas relacionado con la Detección Automática de Plagio, Paráfrasis en idioma español, similaridad de textos, procesamiento de textos en idioma Español, y Reconocimiento de Locutor.*

Situación Problemática: *La detección de similaridad en textos escritos en lenguas naturales es un área de investigación actual. Mayoritariamente los trabajos existentes se realizan sobre el idioma inglés. Los recursos disponibles como corpus o lexicones están personalizados para la lengua inglesa. Dentro de estos recursos uno de los más utilizados es el Wordnet, una base de datos léxica con más de 30 años de desarrollo. Wordnet en inglés posee más de 100 mil términos, mientras que el Wordnet en español solo posee unos 20 mil. Las herramientas para la edición de estos diccionarios son tecnologías privadas de instituciones científicas que solo divulgan el resultado. Se desea aplicar a los algoritmos de detección de similaridad en español similares recursos a los existentes para idioma inglés. Una dificultad con la generalización de estos recursos es que el trabajo de los lingüistas debe hacerse sobre herramientas de fácil acceso y uso.*

Problema científico: *¿Cómo mejorar la detección de similaridad semántica entre palabras en idioma español?.*

Idea a Defender: *La elaboración de una aplicación simple para la edición del diccionario Wordnet en español permitiría la inclusión de términos especializados mejorando el procesamiento de similaridad en palabras del idioma español.*

Objeto de estudio: *Procesamiento del Lenguaje Natural Español.*

Campo de acción: *Herramientas para el procesamiento del lenguaje natural español.*

Objetivo General: Desarrollar un módulo para la herramienta QtNLP que permita agregar palabras al Wordnet en español en la estructura adecuada.

Objetivos Específicos:

- Objetivo I: Sistematizar el estado actual y las tendencias de las herramientas para la edición de wordnet y otros recursos de la computación lingüística.
- Objetivo II: Diseñar la aplicación QtNLP-Wordnet.
- Objetivo III: Implementar la aplicación QtNLP-Wordnet.
- Objetivo IV: Validar la solución implementada.

Tareas de Investigación:

- Revisión de bibliografía actual, y documentación de al menos 3 proyectos que editan recursos léxicos importantes.
- Revisar las tecnologías utilizadas en las aplicaciones de tipo frontend y backend para el desarrollo de recursos lingüísticos.
- Diseñar la interfaz de QtNLP-Wordnet atendiendo a buenas prácticas de experiencia de usuario.
- Diseñar las pruebas para la herramienta diseñada.
- Implementar la interfaz de QtNLP-Wordnet en el QtDesigner.
- Implementar el módulo QtNLP-Wordnet utilizando la arquitectura modular de QtNLP.
- Introducir los 5 primeros términos especializados sobre computación.
- Introducir 15 nuevos términos con un lingüista y evaluar la usabilidad y estabilidad de la aplicación.

Posibles resultados:

- *Interfaz gráfica en Qt y python para la edición del Wordnet en español.*
- Versión editada del diccionario léxico Wordnet en español con 20 nuevos términos sobre computación.

Elaborado por: <Abel Meneses Abad>, <26 de octubre de 2015>