

Concurso de Ideas/Proyectos de Investigación Plataforma MITIC

Informe de evaluación
9 de mayo de 2013

Antecedentes

La Fundación Dr. Manuel Sadosky convocó a grupos de investigación en Recuperación de la Información, Minería y Análisis de Datos, Minería de Textos, Procesamiento de lenguaje, Economía del Conocimiento o áreas afines a presentar Ideas/Proyectos destinados a mejorar la usabilidad, calidad de datos, aprovechamiento, eficiencia o cualquier otro aspecto vinculado a la plataforma MITIC (Mapa de la Innovación en TIC, <http://mitic.fundacionsadosky.org.ar>). La Fundación ofreció financiar la definición, especificación y/o implementación de la o las ideas seleccionadas y su incorporación a la plataforma. El plazo para la recepción de propuestas venció el 4 de mayo de 2013.

Propuestas recibidas

Se recibieron cuatro propuestas de cuatro grupos distintos. A continuación presentamos un resumen de las mismas.

Proponente: Grupo de Investigación-Acción "Sinapsis", dependiente de la Especialización en Diseño de Interacción con Estándares de Accesibilidad y Usabilidad de la Universidad Tecnológica Nacional.

Resumen: La propuesta contiene un plan de investigación-acción para mejorar la usabilidad, el diseño de interacción, y la navegación en general de la interfaz de usuario de la plataforma. Los pasos del plan son la realización de una evaluación heurística sobre los criterios de usabilidad, una investigación de usuarios, la redacción de un informe, el diseño de un prototipo y pruebas de usabilidad sobre el mismo, en dos iteraciones, y la redacción de un informe final con el detalle de las mejoras propuestas.

Proponente: Grupo de Lenguajes, Lingüística e Información (GALLI), Depto. de Computación, FCEyN, UBA

Resumen: La propuesta tiene dos partes. En la primera, se propone el diseño y la implementación de un prototipo de un sistema de question answering (respuesta a preguntas) para el área de las TICs. La idea es analizar la información estructurada en la base de conocimiento de MITIC a fin de utilizar la información semántica disponible, para su utilización en la elaboración de una respuesta. Luego, implementar el sistema de respuesta a preguntas basado en Qanus, utilizando herramientas de reconocimiento de entidades y herramientas de clasificación de preguntas e indexando información complementaria en la web para acceder a respuestas no disponibles en las bases de información locales del MITIC. Proponen desarrollar una interfaz web para el acceso y testeado del sistema. Como entregable final, un informe describiendo el análisis de los resultados del sistema, recomendaciones y propuesta de implementación. La segunda parte de la propuesta, tiene que ver con Calidad de Datos. Aquí, proponen aplicar una metodología para analizar y

mejorar la calidad de los datos de MITIC con actividades que incluyen la determinación de datos críticos, la definición y ejecución de métricas de calidad, y la propuesta de mejoras incluyendo prototipos de algoritmos de mejora de calidad de datos.

Proponente: Grupo de Investigación y Desarrollo en Data Mining y KDD, Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de La Matanza.

Resumen: La propuesta parte de la afirmación de que la base de datos inicial de MITIC posee incompletitudes o incorrecciones, y que la calidad y la dimensión de la base de datos actual restringe el uso de la herramienta impidiendo que se detecten reales mejoras o deficiencias en las facilidades de uso. Proponen para resolver esos problemas "incorporar a su base de datos una carga inicial adecuada". Proponen la utilización de crawlers para recorrer una serie de sitios de universidades, grupos o congresos de TIC.

Proponente: Equipo de PLN, Machinalis (Facultad de Matemática, Astronomía y Física – Universidad Nacional de Córdoba)

Resumen: La propuesta apunta a completar la información asociada a cada uno de los actores de la base de datos, basándose principalmente en métodos de Procesamiento del Lenguaje Natural (PLN) profundos. El componente central de la propuesta son los métodos de Análisis Semántico Latente (Latent Semantic Analysis, LSA), que permiten incorporar semántica a la información, y así unificar y generalizar la información, y a partir de eso encontrar patrones y completar información faltante. Proponen para esto varios pasos: obtener información adicional de algunos actores, procesar toda la información textual asociada a cada actor, identificando términos clave generales de MITIC, armar una representación de cada actor en el espacio de términos clave, y enriquecer este espacio mediante LSA. Para lo anterior se utilizaría análisis lingüístico, métodos estadísticos, análisis semántico latente, y otras herramientas. El objetivo final es enriquecer la base de datos de la Red MITIC con un número significativamente mayor de términos, que tendrán un uso más generalizado a través del universo de actores, eliminando diferencias en el uso de términos para describirse.

Análisis de las propuestas recibidas

El análisis de las propuestas recibidas permitió llegar a las siguientes conclusiones:

Grupo Sinapsis – UTN. El proyecto resulta interesante desde el punto de vista metodológico, en tanto propone aplicar una serie de pasos coherentes con los objetivos de mejorar la usabilidad del sitio, y el equipo que se propone para desarrollarlo posee experiencia y conocimientos en el tema. Por otro lado, el objetivo planteado no está entre los de máxima prioridad para la Fundación Sadosky en este periodo.

Grupo GALLI – UBA. El objetivo de desarrollar un sistema de respuesta a preguntas para el área de las TICs es sumamente interesante pero excede en cierta medida las necesidades de la plataforma MITIC y el alcance del concurso

de ideas. Por otro lado, no resultan del todo claro los resultados esperados de la propuesta presentada, así como las actividades, cronograma y entregables. En lo que respecta a la mejora de la calidad de los datos de MITIC en forma sistemática y metodológica, ésta sí se encuentra entre las necesidades prioritarias del proyecto, y la propuesta resulta interesante. El equipo de investigadores propuesto es, en ambos casos, de mucha confiabilidad.

Grupo de Investigación y Desarrollo en Data Mining y KDD, UNLM. El objetivo de aumentar la cantidad y mejorar la calidad de los datos de la plataforma a partir del crawling de una serie de sitios que se determinen a partir del análisis del campo del conocimiento de las TIC ya forma parte de la estrategia seguida hasta ahora y del plan de trabajo en curso. Si bien algunas de las fuentes sugeridas en la propuesta pueden ser tenidas en cuenta, la misma peca de excesiva simplificación sobre algunos aspectos, como ser la normalización de direcciones o nombres, el descubrimiento de la estructura de las fuentes, etc.

Grupo: Equipo de PLN, Machinalis (FAMAF – UNC)

Análisis: La propuesta resulta de alto atractivo tanto desde el punto de vista teórico o de su grado de innovación, como desde el de su aplicabilidad, utilidad práctica y valor para la plataforma. La propuesta adolece de algunas fallas, en particular resultan poco precisas las descripciones de plazos y entregables de cada etapa. Además, algunas de las ideas propuestas ya fueron consideradas en la versión actual de la plataforma. Los objetivos son ambiciosos pero a la vez realistas y tienen en cuenta la compatibilidad con lo que ya se encuentra desarrollado. El equipo de trabajo propuesto goza de antecedentes muy interesantes en el área del conocimiento involucrada y en aplicaciones similares.

Resultados

A partir del análisis realizado, la Fundación Sadosky toma las siguientes decisiones:

Grupo Sinapsis – UTN

Decisión: Se solicitará modificar la propuesta de trabajo en forma tal de acercarla a las necesidades actuales de la plataforma y a las posibilidades presupuestarias en relación a ellas. En caso de llegarse a un acuerdo, se procederá a financiar el proyecto.

Grupo GALLI – UBA

Decisión: Se propone revisar conjuntamente la propuesta en cuanto finalice la actual colaboración en marcha con el mismo grupo, priorizando los objetivos de calidad de datos y reformulando los vinculados con Question Answering. No se financiará el proyecto por el momento.

Grupo de Investigación y Desarrollo en Data Mining y KDD, UNLM.

Decisión: No se considera prioritaria su financiación y no se le dará curso por este año.

Grupo: Equipo de PLN, Machinalis (FAMAF – UNC)

Decisión: Se iniciarán conversaciones para modificar los aspectos deficientes de la propuesta. Un aspecto formal a modificar tiene que ver con la personería del grupo, que de acuerdo a las condiciones del llamado debe pertenecer a Universidades o Institutos de Investigación nacionales. En caso de llegarse a un acuerdo, se procederá a financiar el proyecto.