



PROYECTO: Metodología para determinar la exactitud de una respuesta, escrita en forma textual, a un interrogante sobre un tema específico, aplicando herramientas informáticas.

Resumen Técnico

El presente proyecto realizará análisis de texto en un párrafo, de redacción libre, que es respuesta a preguntas relacionadas a un dominio específico, con el objetivo de detectar si esa respuesta es correcta. El ámbito de prueba se realizará usando como sujetos de prueba a los alumnos de una determinada cátedra, que responderán a preguntas de exámenes usando texto de redacción libre. La cátedra elegida es Paradigmas de Programación del 2do año de la carrera Ing. en Sistemas de Información en UTN FRC, ya que la misma posee contenidos con dominios simples y directos que facilitan la interpretación de las posibles respuestas de los alumnos. Para este análisis se definirá el dominio de aplicación tanto en relación a la temática que se trate como en la forma en que el alumno deba dar las respuestas. Se estudiará la necesidad de establecer para las respuestas ciertas restricciones que facilitarían su posterior estudio. Para realizar el análisis del texto, se comenzará con el análisis ortográfico de cada palabra (ya que es necesario que la palabra esté correctamente escrita a fin que el sistema la comprenda) y se obtendrá una segunda versión de la oración sin errores ortográficos, así será posible saber si las palabras pertenecen o no al dominio de trabajo. A continuación se evaluará la necesidad de realizar un análisis sintáctico de las oraciones, en tal caso se obtendría una tercera versión de la respuesta a ser analizada. Sabiendo que las palabras están escritas correctamente, se evaluará la aplicación de la técnica de grafos tanto para conocer si las oraciones pertenecen al dominio en cuestión, si están construidas correctamente y que correspondan a su respuesta correcta, así como para explicar el dominio de trabajo. Es importante indicar que el proyecto presente no pretende profundizar en el conocimiento de la minería de textos, ya que en base a la recuperación de la información desde el texto escrito por el alumno, se tratará de identificar cual es la respuesta provista, y si es correcta, en comparación con todas las posibles respuestas a una pregunta dentro del dominio de trabajo. En cambio, text mining es el análisis de recursos de textos que son nuevos, en los que información previamente desconocida es descubierta. Éste es un campo interdisciplinario que toma técnicas del campo general de data mining y, adicionalmente, combina metodologías de varias otras áreas tales como extracción de información (Information Extraction - IE), recupero de información (Information retrieval - IR), lingüística computacional, categorización, temas de seguimiento y concepto de vínculo. Es frecuente confundir el campo de Information Retrieval (IR) y el de text mining, véase [6] y [8]. Esto sucede porque ambos tratan con el texto y sus particularidades, entonces ambos tienen cuestiones de apariencia similar. IR ha dejado muchos algoritmos y métodos a text mining. La diferencia entre estos 2 campos es principalmente su acierto final. En IR, el objetivo es recuperar documentos que parcialmente concuerdan con una consulta y seleccionar de esos documentos algunos de los mejores aciertos. Text mining trata acerca de descubrir hechos desconocidos y verdades ocultas que deben existir en el léxico, la semántica o relación estadística de las colecciones de textos. Otro campo que ha dejado metodologías a text mining es Extracción de Información (IE). IE difiere de text mining en que extrae datos estructurados específicos (Ej. nombre de personas, ciudades, títulos de libros) y relaciones re-especificadas tal que se descubren nuevas relaciones y generan nuevos diseños. En text mining la información encontrada es no sospechada, en cambio en IE es predefinida y se concuerda con un interés especificado por el usuario. Las técnicas de IE pueden ser parte de las tareas de text mining con el fin de facilitar la extracción de conocimiento.

Autores:

Grosso, Mario Alberto
Marciszack, Marcelo Martin
Paz Menvielle, María Alejandra
Liggorria, Karina; Guzman, Analía; Casatti, Martín Gustavo; Higa Tamashiro, Seiyu Ricardo; Gimenez, Juan Pablo

Duración: Inicio: 01/01/2015 - Fin: 31/12/2017