

El uso de NLP e IA para mejorar la calidad y extraer información de especificaciones en lenguaje natural

Marcela Samela¹, Gustavo Ernesto Gonzalez¹, Mariana Otero¹, Lovena Laurent¹, Guillermo Omar Pizarro Vasquez^{1,2}, Adrian Federico Gil¹, Dario Cardacci¹, Juan Dalotto¹, Leandro Antonelli^{1,3}

¹CAETI, Facultad de Tecnología Informática, Universidad Abierta Interamericana, CABA, Argentina.

²Universidad Politécnica Salesiana, Ecuador.

³Lifia, Facultad. de Informática, Universidad Nacional de La Plata, Calle 50 y 120 - La Plata - Bs. As. – Argentina.

<mailto:{marcela.samela,gustavoernesto.gonzalez,mariana.otero,lovena.laurent,guillermoomar.pizarrovasquez,adrianfederico.gil,dario.cardacci,juaneduardo.dalotto,leandro.antonelli}@uai.edu.ar>

CONTEXTO

El proyecto de I/D que se presenta en este artículo lo realiza un grupo de docentes e investigadores de la Universidad Abierta Interamericana vinculados a un proyecto de investigación del centro CAETI de la misma Universidad.

INTRODUCCIÓN

Se conoce como revolución industrial al proceso en el cual con la invención de la máquina a vapor se lograron automatizar procesos en los que intervenía la fuerza humana, logrando resultados más rápidos y económicos. Algunos autores identifican 4 revoluciones industriales, en donde la tercera y cuarta etapa dependen fuertemente de los sistemas informáticos.

Lo cierto es que actualmente los sistemas informáticos están en todos los aspectos de nuestra vida, y la cantidad de información que se produce y consume es muy grande. Hay información que está codificada y que sólo entienden los sistemas informáticos, pero también hay mucha información escrita en lenguaje natural que generan y consumen los humanos.

Con el fin de optimizar el tratamiento de esta información en lenguaje natural, este proyecto tiene por objetivo desarrollar y evaluar herramientas basadas en Procesamiento de Lenguaje Natural e Inteligencia Artificial para el tratamiento de especificaciones técnicas y no técnicas con el fin de mejorar la calidad, extraer información y consolidar el conocimiento.



Veo a un hombre en un barco con un telescopio

LINEAS DE INVESTIGACION Y DESARROLLO

Este proyecto se lleva a cabo a través de un método iterativo e incremental, el cual comienza con un relevamiento del estado del arte, para seleccionar una o varias técnicas, y proponer una contribución al estado del arte en algún área o dominio específico.

A través de diferentes iniciativas independientes se busca explorar soluciones en diferentes dominios para evaluar y analizar la posibilidad de llevar la contribución a un dominio diferente, con la eventual generalización de la contribución.

Básicamente, el proyecto buscará analizar estilos de redacción con el fin de sugerir mejoras a la vez de estandarizar la estructura. De esta forma, será más sencillo el comparar distintos textos para establecer similitudes. Con esto se podrán agrupar textos relacionados y poder extraer información de los mismos. Todas estas acciones apuntan a consolidar la información permitiendo que sea procesada de una forma más eficiente.

RESULTADOS ESPERADOS

El objetivo general del proyecto es el de diseñar, implementar y testear herramientas basadas en NLP e IA para el tratamiento de documentos (especificaciones) técnicas y no técnicas, con el fin de mejorar la calidad y consolidar la información que proveen.

Los objetivos específicos se basan en explorar distintos matices: (i) analizar la calidad de las especificaciones con el fin de sugerir mejoras, (ii) agrupar y categorizar documentos, (iii) comprar nivel de similitud, (iv) identificar conceptos relevantes y sus relaciones, y (v) consolidar la información extraída con el fin de darle un tratamiento automatizado.

FORMACIÓN DE RECURSOS HUMANOS

Se espera formar a jóvenes docentes investigadores, introducir a recientes graduados en el mundo académico de la investigación y finalmente a brindar a estudiantes avanzados la posibilidad de que un marco para llevar un trabajo final de carrera.