

Uso de Natural Language Processing y Machine Learning para la Gestión de Requerimientos

Juliana Delle Ville¹, Giuliana Maltempo¹, Roxana Portugal^{2,3}, Leandro Antonelli^{1,4}

¹Lifia, Fac. de Informática, UNLP, La Plata, Bs As, Argentina

²Universidad Nacional de San Antonio Abad del Cusco, Cusco, Perú

³Ludwig Maximilians Universität, München, Deutschland

⁴CAETI - Facultad de Tecnología Informática - Universidad Abierta Interamericana

{juliana.delleville, giuliana.maltempo,
leandro.antonelli}@lifia.info.unlp.edu.ar
roxana.quintanilla@unsaac.edu.pe

Abstract— La ingeniería de requerimientos es una de las etapas más críticas del desarrollo de software. Si bien existen desarrollos bajo métodos ágiles en donde las prácticas de requerimientos son mínimas, aún existen muchos desarrollos tradicionales en donde se invierte mucho esfuerzo en requerimientos. En particular, el lenguaje natural es el instrumento más utilizado, puesto que ayuda a reducir la brecha comunicacional que existe entre los clientes y el equipo de desarrollo. Sin embargo, el lenguaje natural puede originar varios defectos: ambigüedad, vaguedad, etc. En últimos años dos disciplinas de la inteligencia artificial han contribuido con muchas herramientas que han revolucionado el tratamiento de texto: natural language processing y machine learning. Este tutorial tiene por objetivo brindar las herramientas básicas para aplicar en la ingeniería de requerimientos.

Keywords- procesamiento de lenguaje natural, inteligencia artificial, gestión de requerimientos.

1. Introducción

La ingeniería de requerimientos es una de las etapas más críticas en el desarrollo del software. Los errores cometidos en esta etapa, pueden costar hasta 200 veces repararlos en etapas posteriores. Si bien los métodos ágiles intentan mitigar este problema, aún se desarrollan muchos sistemas de software con modelos tradicionales, en donde hay un gran volumen de especificaciones, las cuales son generalmente en lenguaje natural. Sin embargo, el lenguaje natural puede originar varios defectos [2]: ambigüedad, vaguedad, etc. Por lo tanto, es importante contar con herramientas que ayuden a automatizar el proceso de análisis de las mismas. Natural language processing y machine learning [1] [4] [5] proveen herramientas para: (i) analizar texto, identificar patrones y poder extraer ciertas palabras claves según su valor gramatical (sustantivos, verbos, etc.), (ii) identificar entidades (nombre propios,

locaciones geográficas, magnitudes, etc.), (iii) comparación de similitud, (iv) categorización, y (v) desambiguación entre otras herramientas. Este tutorial tiene por objetivo brindar las herramientas básicas de natural language processing y machine learning que podrá ser utilizado para la gestión de requerimientos. Si bien un algoritmo no puede reemplazar al analista funcional, al menos puede automatizar ciertas tareas para reducir el esfuerzo. El resto de la propuesta se organiza de la siguiente manera. La sección 2 describe la audiencia esperada de este tutorial. La sección 3 describe los temas del tutorial.

2. Audiencia, dinámica del tutorial y temario

Este tutorial está destinado a personas con conocimiento en programación. El tutorial será orientado a ejemplos concretos de aplicación. En el mismo se darán los conceptos teóricos y se brindarán ejemplos de aplicación (código fuente). También se brindarán ejercicios para que los participantes puedan realizar luego de finalizado el tutorial. De esta forma, luego del curso, los participantes podrán automatizar ciertas tareas con sus propios documentos. Los temas que se tratarán en el tutorial son los siguientes [3]:

- Natural Language for Requirements Engineering (NL4RE)
- Introducción al text mining. Lenguaje Natural.
- Explotación de información vía reglas. Spacy pipeline. Rule-based Matcher. Entities. Dependency Matcher
- Explotación de información vía algoritmos de Machine Learning.
- Machine Learning. Modelos supervisados y no supervisados. Error cuadrático medio. Descenso del gradiente.
- Preprocesamiento. Normalización. Stopwords. Stemming. Lemmatizing.
- Algunas herramientas. Named Entity Recognition Comparación de Similitud. Categorización. Word desambiguation. Modelado de tópicos

Referencias

1. Al-Safadi, L. A. E.: "Natural Language Processing for Conceptual Modeling." J. Digit. Content Technol. its Appl. 3, pp 47-59, 2009.
2. Berry, D., Kamsties, E. and M. Krieger. From Contract Drafting to Software Specification: Linguistic Sources of Ambiguity. Handbook, University of Waterloo, 2003.
3. Kulkarni, A., Shivananda, A.: "Natural language processing recipes: Unlocking text data with machine learning and deep learning using python" APress, 2a ed. (2021).
4. Robeer, M., Lucassen, G., van der Werf, J. M. E. M., Dalpiaz, F., Brinkkemper, S.: "Automated Extraction of Conceptual Models from User Stories via NLP," 2016 IEEE 24th International Requirements Engineering Conference (RE), Beijing, China, pp. 196-205, 2016.
5. Shuttleworth, D. Padilla, J.: "From Narratives to Conceptual Models via Natural Language Processing," 2022 Winter Simulation Conference (WSC), Singapore, pp. 2222-2233, 2022.