

GENERACIÓN AUTOMÁTICA DE CÓDIGO SOFTWARE A TRAVÉS DE MODELOS PREENTRENADOS DE LENGUAJE A PARTIR DEL IDIOMA ESPAÑOL

AUTOMATIC CODE GENERATION FROM SPANISH LANGUAGE USING PRETRAINED LANGUAGE MODELS

Investigadores USAL:

Bender, Adrián Simón (bender.adrian@usal.edu.ar); Nicolet, Santiago Javier; López, Juan José; Folino, Pablo Daniel

Alumno USAL:

Hansen, Gustavo

Palabras clave: generación de código, GPT, automatización

Keywords: *code generation, GPT, automation*

Resumen

El aprendizaje automático ha generado importantes logros en los últimos años, y cada vez más disciplinas se apoyan en él permitiendo, entre otras cosas, automatizar tareas que requieren conocimiento y experiencia de profesionales. En el último tiempo se ha comenzado a utilizar aprendizaje de máquinas en el área del desarrollo de *software* y así, por ejemplo, está surgiendo una nueva generación de herramientas que, entre otras cosas, permite escribir un código más confiable y detectar errores en forma automática. Si bien estas herramientas son utilizadas actualmente para complementar la labor de los profesionales, el avance que ha tenido la IA en otros campos nos hace pensar en un mayor grado de automatización en tareas relativas al desarrollo de *software*. Los modelos preentrenados de lenguaje (como GPT), a través de sus últimos avances, están demostrando importantes casos de éxito en tareas que implican generación de textos a partir de un estímulo escrito en lenguaje natural. Este proyecto pretende explorar el empleo de estos modelos para la generación de código de *software* en busca de determinar su aplicabilidad, conocer la calidad del código generado y poder vislumbrar algunas fortalezas y debilidades de esta tecnología para la generación automática de *software*, sobre la base de una especificación escrita en idioma español.

Abstract

Machine Learning has generated important achievements in recent years. More and more disciplines rely on it, allowing, among other things, to automate tasks that require knowledge and qualified expertise. Recently, Machine Learning has begun to be used in the area of software development. Thus, for example, a new generation of tools has emerged that, among other things, allows more reliable code to be written and errors to be detected automatically. Although these tools are currently used to complement the work of professionals, the progress that AI has had in other fields makes

us think of a greater degree of automation in tasks related to Software Development. Pretrained Language Models (such as GPT), thanks to their latest updates, are showing important success cases in tasks that involve generating texts from an input written in natural language. This project aims to explore the use of these models for the code generation with the goal of establishing its applicability, knowing the quality of the generated code, and being able to perceive some strengths and weaknesses of this technology for the automatic code generation based on a specification written in Spanish language.