

Empirical Analysis on OpenAPI Topic Exploration and

Discovery to Support the Developer Community¹

Leonardo da Rocha Araujo¹, Guillermo Rodríguez¹, Santiago Vidal¹

Claudia Marcos^{1,2} and Rodrigo Pereira dos Santos³

¹*ISISTAN Research Institute (CONICET-UNICEN), Tandil, Buenos Aires, Argentina.*

³*UNIRIO, Rio de Janeiro, Brazil*

Abstract. OpenAPI has become a dominant standard for documentation in the service-oriented software industry. OpenAPI is used in many analysis and reengineering approaches for RESTful service and microservice-based systems. An OpenAPI document has several components that are usually filled by humans using natural language (e.g. description of a certain functionality). Thus, subjectivity may lead to inconsistencies and ambiguities. Understanding what an API does is a challenging question. As a consequence, this issue could hinder developers from identifying the functionality of APIs, after reading all its components. Along this line, we argue that developers will be provided with supportive tools to find those APIs that better suit their needs. In this paper, we propose a step towards creating these kinds of tools by empirically analyzing a set

¹ da Rocha Araujo, L., Rodríguez, G., Vidal, S., Marcos, C., & dos Santos, R. P. (2021). Empirical Analysis on OpenAPI Topic Exploration and Discovery to Support the Developer Community. *Computing and Informatics*, 40(6), 1345-1369.

of 2,000 OpenAPI documents with the goal of extracting the main topics of an API using three topic modeling algorithms.

To address this issue, we focus on three tasks: i) determine which component of an OpenAPI document provides the most meaningful information, ii) compare three state-of-the-art topic modeling algorithms, and iii) determine the optimal number of topics to represent an API. Our findings show that the best results could be obtained from the Description component by using the Non-negative Matrix Factorization (NMF) or Latent Semantic Indexing (LSI) algorithms. To help developers find services in the OpenAPI directory, we also propose a prototype tool to explore the OpenAPI documents and analyze extracted topics to assess if the APIs meet developer's needs.

Keywords: Microservices, OpenAPI, Migration, Legacy Systems, Topic Modeling.