Analyzing the Natural Language Processing technology field using Tech mining

Gaizka Garechana, Rosa Río-Belver, Izaskun Álvarez-Meaza, Enara Zarrabeitia
Business Management Department, University of the Basque Country (UPV/EHU), Spain.

Abstract

The Natural Language Processing (NLP) field is the branch of computational science devoted to the automated interpretation of human language, having several technical applications in areas such as speech recognition and information retrieval/summarization, among others. In this paper we analyze NLP patent data corresponding to the yearly interval 2006-2020 in order to characterize the main agents, purposes and analytical tools behind this field. With this goal in mind, we use text mining software to extract the relevant information from patent abstracts and identify the specialization of the main players in the area. In addition to this, we detect the dominance of artificial intelligence applications of NLP and the versatility and acceptation of deep learning algorithms in this field. These concepts are at the same time the dominant ones and show the highest growth rate, being present in roughly 15% of the patents forming our dataset. Two clear conclusions are extracted when analyzing the conceptual maps and cluster analysis of the data: voice/speech recognition systems and the automated medical diagnosis systems are well-consolidated specialties in NLP patenting activity.

Keywords: Tech mining; NLP; Patent.