

Title	ニューラルネットワークに基づいたセマンティック分析による特許請求の分割
Author(s)	Silva De Carvalho, Danilo
Citation	
Issue Date	2018-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/15318
Rights	
Description	Supervisor:NGUYEN, Minh Le, 情報科学研究科, 博士

Abstract

The transition from the last century brought a considerable number of changes to international relations and trade, with a sharp acceleration of the Globalization process. The previous focus of industrial development is being changed from the production of physical assets to the concept of Intellectual Property, which is regulated in most countries by the patent system. With an increasing number of patents being applied and granted, management of innovation related information became an arduous task, leading to the development of a variety of approaches for its automation. While the use of Natural Language Processing techniques is well established among such approaches, the characteristics of this type of document still provides challenges to be addressed.

In this dissertation, I present a method for identifying and classifying textual segments from patent documents into relevant information types. It aims to facilitate the categorization and comparison of inventions by means of semantic analysis and is centered on claim sentences. The claims are the main information source in patent reviewing and litigation, so they are an important target for automation. The central aspects of the presented method, and major contributions of this work, are the annotation methodology for identifying elements of ideas in patent claims, that emulates the workflow of a patent professional, and the gathering and exploitation of multiple linguistic features from different sources. Those features enable the use of powerful Machine Learning techniques known as Deep Learning (i.e., Deep Artificial Neural Networks) even in cases where available training data is scarce. This research comprises the study of several aspects of patent information processing and describes the development of novel approaches to deal with them. In particular, the structuring of patent documents and claims from digitalized paper forms; the computational representation of lexical and semantic aspects of natural language; and the development and optimization of Deep Learning architectures for claim annotation. The analysis of claims is grounded on linguistic concepts and segmentation is done in a principled way, following patent expert advice. A manually annotated dataset for patent claims complete the set of contributions. Some basic premises explored in this work regard the syntactic constraints found in patent claims, which allow simplifications in appropriate methods or the use of techniques that would otherwise not be effective, while avoiding the otherwise effective solutions that fail when applied to claims. Experimental evaluation is performed for each one of the solutions presented, and the benefits and drawbacks of the developed methods discussed in detail. Results indicate that the automated segmentation of claims using the proposed method is viable and produces the desired results. The annotation principles guarantee the usefulness of the results to the patent expert community.

Keywords: Claim segmentation, Patent Information Processing, Semantic analysis, Deep Learning, Natural Language Processing