PAT DIV: A CHRONOLOGICAL PATENT DIVISIONAL APPROACH

[1]Dr. R. Palson Kennedy, [2]V.Karthik, [3]Dr. Kalyana Sundram

[1] Professor , PERI Institute of Technology, [2]Research Scalar, Anna University, Chennai, [3]Asso Prof , Shaveetha University.

ABSTRACT

Copy right s plays an important role in industry as well as in academic institutes. However, quickly understanding the real meaning of a given copy right or a Pattern right is difficult as they typically are very long and written in a quite unreachable way. These important information, especially the innovation itself and the experimental part of the invention, are usually contained in the description section. However, in many Pattern right s the explanation parts are neither interpreted nor easily demonstrable. Here, we explain our novel PatDiv method for Pattern right divisions, which aims at mechanically and openly recognize the most important parts of a Pattern right . PatDiv uses a two-step approach, where a Pattern right is first divisional into text blocks in an unsupervised fashion followed by a supervised classification step for each identified division. In contrast to previous work, PatDiv uses semantic word embeddings in both phases and applies a chronological learning algorithm for the second step. These changes guide to, on average, an improvement of 9.37% (8.87%, 9.00%) in terms of F1-score (precision, recall) and 7.19 in terms of accuracy in association to a baseline, as evaluated on two novel and manually divisional gold standard Pattern right corpora. The method also is easily parallelizable, fast, making it applicable for truly large Pattern right collections.

Keywords: Pattern right divisions; Machine learning; Text divisions; Chronological classifiers; Word embeddings.

1. INTRODUCTION

Pattern right s are a key source of information for most industries. They are often the first (and sometimes the only) channel of publication of new ideas, innovations, and technologies. A big advantage of Pattern right s is that they have to describe their innovative aspects in detail, because international Pattern right laws demand full explanation of the invention to receive Pattern right protection. Accordingly, scientists, Pattern right searchers, data analysts, and Pattern right lawyers are continuously reading and analyzing Pattern right documents to learn about novel trends, to detect Pattern right infringement, to valuate a competitor's intellectual property, to assess own novel ideas etc.

Due to the steep growth of the number of published Pattern right applications and due to the typical length of Pattern right applications, it has become extremely challenging to keep track of novel innovations even when focusing on a single technical field [1], [2]. One possible remedy is the application of computational methods from the field of information extraction (IE) to automatically extract relevant knowledge from Pattern right s [3], [4], [5], [6], [7]. However, when applied to entire Pattern right s, IE methods often produce an enormous amount of information which is mostly irrelevant for typical analysis tasks, such as prior art, extensive lists of slightest variations in potential usages of the innovation, lengthy metaphors of the inner workings of the invention, etc. Therefore, a typical and important pre-processing step before applying IE is the identification of the most relevant parts of a Pattern right .

Defining and finding the most relevant parts of a Pattern right is not a trivial matter. Typically, Pattern right s are pre-structured in sections such as bibliographic data, title, abstract, description, or claims. In these, title and abstract often