

## PaSA: A Dataset for Patent Sentiment Analysis to Highlight Patent Paragraphs

**Authors :** Renukswamy Chikkamath, Vishvapalsinhji Ramsinh Parmar, Christoph Hewel, Markus Endres

**Abstract :** Given a patent document, identifying distinct semantic annotations is an interesting research aspect. Text annotation helps the patent practitioners such as examiners and patent attorneys to quickly identify the key arguments of any invention, successively providing a timely marking of a patent text. In the process of manual patent analysis, to attain better readability, recognising the semantic information by marking paragraphs is in practice. This semantic annotation process is laborious and time-consuming. To alleviate such a problem, we proposed a dataset to train machine learning algorithms to automate the highlighting process. The contributions of this work are: i) we developed a multi-class dataset of size 150k samples by traversing USPTO patents over a decade, ii) articulated statistics and distributions of data using imperative exploratory data analysis, iii) baseline Machine Learning models are developed to utilize the dataset to address patent paragraph highlighting task, and iv) future path to extend this work using Deep Learning and domain-specific pre-trained language models to develop a tool to highlight is provided. This work assists patent practitioners in highlighting semantic information automatically and aids in creating a sustainable and efficient patent analysis using the aptitude of machine learning.

**Keywords :** machine learning, patents, patent sentiment analysis, patent information retrieval

**Conference Title :** ICAIL 2022 : International Conference on Artificial Intelligence for Law

**Conference Location :** Singapore, Singapore

**Conference Dates :** May 05-06, 2022