"Automated Research Paper Classification and Retrieval Using .NET and NLP"

Automated classification and retrieval of research papers are critical tasks in the realm of academic research and knowledge discovery. With the increasing volume of academic publications, it has become challenging to manually sift through papers to identify relevant ones. This paper introduces an automated research paper classification and retrieval system using the .NET platform and natural language processing (NLP) techniques. The system is designed to classify research papers based on their abstract and retrieve papers that match user queries. First, the system extracts the abstract text from research papers, stored in PDF format, using the iTextSharp library for text extraction. Once the text is retrieved, several NLP techniques are applied to preprocess the text, including stopword removal, stemming, and tokenization. For the classification task, a supervised machine learning approach is employed, where labeled datasets are used to train classifiers. Algorithms such as Naive Bayes, Support Vector Machines (SVM), and Random Forests are implemented to classify the papers into predefined categories such as "Artificial Intelligence," "Data Science," "Bioinformatics," and "Machine Learning." Additionally, for retrieval purposes, the system uses vector space models and similarity measures to compare the user's query with the abstract texts of the research papers. The retrieval process is enhanced by applying semantic analysis methods such as Latent Semantic Indexing (LSI) and Word2Vec to capture the contextual meaning of words and phrases within the abstracts. The system's performance is evaluated using several metrics, including precision, recall, and F1-score, to ensure that it effectively classifies and retrieves relevant papers. The results demonstrate that the system can automate the process of research paper classification and retrieval, making it easier for researchers to access the most pertinent papers in a fraction of the time compared to manual searching.