"Implementing a Scalable Abstract Processing Engine with .NET and Azure Cognitive Services"

In academic research, the ability to efficiently process and analyze large volumes of research abstracts is crucial. This paper explores the implementation of a scalable abstract processing engine using the .NET framework integrated with Azure Cognitive Services. The proposed engine leverages the cloud capabilities of Azure to ensure that the processing of research abstracts can be scaled dynamically based on demand. The system uses .NET Core as the backend framework to handle the business logic and interactions with the Azure Cognitive Services API, which provides powerful NLP tools like Text Analytics for sentiment analysis, key phrase extraction, and language detection. The engine begins by extracting text from research papers in PDF format using iTextSharp, followed by preprocessing techniques such as tokenization, lemmatization, and stopword removal. The abstracts are then analyzed using Azure's cognitive services to detect entities, categorize the abstracts, and extract important concepts. To support scalability, the system uses Azure Functions and Azure Logic Apps for cloud-based automation, allowing the processing of abstracts to be distributed across multiple nodes. Azure Blob Storage is used to store the extracted abstracts, and the Azure SQL Database stores metadata such as categories, keywords, and relevant tags. The abstract processing engine is designed to support high throughput and low-latency querying. Users can query abstracts via a web-based interface built using ASP.NET Core, and the system returns relevant abstracts based on the semantic content of the query. Performance evaluations demonstrate the engine's ability to process large datasets efficiently, providing researchers with a robust tool for rapid literature analysis.