Sample PDF Document 1

This is a simple test PDF document for the RAG Question Answering system.

Introduction to Natural Language Processing

Natural Language Processing (NLP) is a field of artificial intelligence that focuses on the interaction between computers and human language. It enables machines to understand, interpret, and generate human language in a valuable way.

Key Concepts in NLP:

1. Tokenization: Breaking text into smaller units called tokens 2. Named Entity Recognition: Identifying entities like names, locations, organizations 3. Sentiment Analysis: Determining the emotional tone of text 4. Machine Translation: Translating text from one language to another 5. Question Answering: Building systems that can answer questions about text

Applications of NLP:

- Chatbots and virtual assistants - Search engines and information retrieval - Text summarization and generation - Language translation services - Voice recognition systems

Retrieval-Augmented Generation (RAG):

RAG is a technique that combines information retrieval with text generation. It helps reduce hallucinations in language models by grounding responses in retrieved factual information from a knowledge base.

The RAG pipeline consists of: 1. Document ingestion and chunking 2. Embedding generation 3. Vector storage and indexing 4. Query processing and retrieval 5. Context-aware answer generation

This approach is particularly useful for building question-answering systems that need to provide accurate, factual responses based on specific documents or knowledge bases.