Alice RAG Test Document

# Dokumentuppladdning & RAG-system

Alice kan ladda upp och analysera dokument i flera format:

* Text-filer (.txt) - Enkel textanalys
* Markdown (.md) - Strukturerad dokumentation
* PDF-filer - Avancerad textextrahering
* Word-dokument (.docx) - Rik textformatering
* HTML-filer - Webbdokument och strukturerad markup

# Semantisk Sökning

Alice använder OpenAI text-embedding-3-small för att skapa vektorrepresentationer av text. Detta möjliggör semantisk matchning där Alice kan hitta relevant information även om de exakta orden inte matchar.

Exempel:

Fråga: "När slutar skolan på måndagar?"

Hittar: "Måndag - Skoldag slutar 14:00"

Trots att orden inte matchar exakt förstår Alice intentionen.

# Chunk-strategier

För stora dokument använder Alice intelligent chunking:

• Maximalt 1000 tecken per chunk

• Bevarar meningsstrukturer

• Skapar separata embeddings för varje chunk

• Länkar chunks till ursprungsdokument

# Testfrågor för detta dokument

1. Vilka filformat kan Alice hantera?

2. Vad är semantisk sökning?

3. Hur stora chunks använder Alice?

4. Vilken embedding-modell används?