**Fundamentals of Knowledge Mining and Azure AI Search**

**Azure AI Search – AI-drivna söklösningar**

**Azure AI Search** är en **Platform as a Service (PaaS)**-lösning som gör det möjligt att skapa avancerade sökapplikationer. Den kan bearbeta data från **structured, semi-structured och unstructured** källor och använda **AI enrichment** för att analysera text och bilder. Eftersom tjänsten är **cloud-based**, hanteras infrastruktur och skalbarhet av Microsoft, vilket gör den enkel att implementera utan att behöva hantera hårdvara.

**Funktioner i Azure AI Search**

Azure AI Search bygger på **Apache Lucene** och erbjuder kraftfulla funktioner för sökning och analys, inklusive:

* **Data ingestion från olika källor** – Tar emot data i **JSON format** och kan automatiskt hämta data från **Azure Storage, Azure SQL Database, Azure Cosmos DB** m.fl.
* **Multiple search options** – Stöd för **vector search, full text search och hybrid search**.
* **AI enrichment** – Integrerad analys av text och bilder via **Azure AI Services**.
* **Linguistic analysis** – Stöd för **56 språk**, inklusive **phonetic matching** och **natural language processing**.
* **Configurable user experience** – Stöd för **vector queries, fuzzy search, autocomplete, geo-search**, och andra avancerade funktioner.
* **Azure integration** – Fungerar med **Azure AI Services, Azure OpenAI**, och andra molntjänster.

**Hur en söklösning byggs i Azure AI Search**

Sökningen i **Azure AI Search** hanteras genom en **indexing pipeline**, där data går igenom flera steg innan det blir sökbart:

1. **Data source** – Ursprunglig lagring av data (t.ex. **Azure Storage, SQL Database** eller **Cosmos DB**).
2. **Indexer** – Automatiserar dataflödet från **data source** och omvandlar informationen till **JSON format**.
3. **Document cracking** – Extraherar textinnehåll från filer (t.ex. **PDFs, bilder, ljudfiler**).
4. **AI enrichment** – Förbättrar datan genom att lägga till information via **OCR, sentiment analysis, entity recognition** m.m.
5. **Push to index** – Serialiserad **JSON data** lagras i ett **search index**.

När ett **search query** skickas in, letar sökmotorn i **search index** och returnerar resultat baserat på indexets struktur.

**Prestanda och skalbarhet i Azure AI Search**

Ett **search index** fungerar som en tabell där data är organiserad enligt en **index schema**, som definierar **fields, data types, attributes** och regler för **filtering och sorting**.

**Index schema** kan innehålla:

* **Text fields** – Sökbart innehåll.
* **Key phrases och entities** – Automatiskt extraherade nyckelord.
* **Image tags** – Tolkade metadata från bilder via **computer vision**.

**Skapa ett search index i Azure**

För att använda **Azure AI Search** måste en **search resource** skapas i **Azure Portal**. Det kan göras genom:

* **Azure Portal Import Data Wizard** – Automatiserad process för att skapa **data source, index, indexer och skillset**.
* **REST API** – För programmatisk hantering av **index creation** och dataingestion.
* **SDKs** – Stöd för integration i **.NET, Python, Java och JavaScript**.

**Import Data Wizard** skapar och kör automatiskt en **indexer** och bygger ett **search index** från de angivna datakällorna.

**AI Enrichment och Knowledge Store**

**AI enrichment** används för att analysera och berika data innan den indexeras. Detta görs genom en **skillset**, som innehåller **built-in skills** såsom:

* **OCR (Optical Character Recognition)** – Extraherar text från bilder.
* **Text translation** – Översätter text till andra språk.
* **Sentiment analysis** – Analyserar textens tonläge (positiv, neutral eller negativ).
* **Entity recognition** – Identifierar personer, platser och organisationer i text.

För att lagra berikad data kan en **Knowledge Store** skapas, där information sparas i **Azure Storage** (tabeller eller blob storage) för vidare analys.

**Query Design i Azure AI Search**

När **search index** är byggt kan det användas för att köra **search queries** via **REST API** eller **SDKs**. En sökfråga returneras som **JSON response** och kan konfigureras för att:

* Definiera vilka **fields** som ska sökas i och returneras.
* Använda **filtering, sorting och scoring** för att anpassa resultaten.
* Begränsa sökresultat baserat på **geo-search** eller andra parametrar.

Det finns två huvudsakliga **query syntaxes**:

1. **Simple syntax** – Grundläggande sökfrågor med stöd för **fuzzy search, autocomplete** m.m.
2. **Full Lucene syntax** – Mer avancerad syntax för komplexa frågor.

Exempel på en **simple query**:

coffee (-"busy" + "wifi")

Denna fråga söker efter *coffee*, men exkluderar *busy* och inkluderar *wifi*.

**Sammanfattning**

**Azure AI Search** är en molnbaserad söktjänst med stöd för **vector search, AI enrichment, text och bildanalys**. Genom en **indexing pipeline** kan data från olika källor processas och göras sökbar. **AI enrichment** möjliggör smarta funktioner som **OCR, sentiment analysis och entity recognition**, vilket förbättrar sökresultaten.

Azure AI Search kan enkelt integreras via **REST API, SDKs eller Azure Portal**, och sökfrågor kan utföras med **simple query syntax eller Lucene syntax**. Kombinationen av **hög tillgänglighet, säkerhet och AI-förstärkt sökning** gör Azure AI Search till ett kraftfullt verktyg för att bygga intelligenta sökapplikationer.