

PROJEKT

KTO: **Lingaro**

INTRO: *Lingaro jest dynamiczną firmą, której misją jest pomoc firmom zdobyć przewagę konkurencyjną poprzez szybsze, lepsze i mądrzejsze decyzje w oparciu o Dane i Analitykę. Unikatowe wartości dostarczane przez firmę zbudowane są na łączeniu zwinności i innowacyjności start-up'u z dojrzałością i doświadczeniem w pracy dla globalnych klientów z listy Fortune 500. Firma obecnie zatrudnia ponad 800 specjalistów w obszarach BI, Big Data, AI & ML, Data Analytics itp..*

CO: **Serverless real-time document similarity engine using Python and Azure Functions**

OPIS: **Istotę projektu stanowi zbudowanie w pełni funkcjonalnego serwisu, rekomendującego artykuły Wikipedii w oparciu o ich podobieństwo do zadanego wyszukiwania. Rozwiązanie oparte na wektorowej reprezentacji dokumentów zrealizowane będzie przy użyciu języka Python. Wdrożenie serwisu zrealizowane zostanie w oparciu o dwa główne komponenty platformy chmurowej Azure: Azure Functions oraz Azure Storage.**

Uczestnicy projektu zrealizują całościowy proces prowadzący do powstania serwisu: zdobędą dane, używając scrapingu zgodnego z polityką Wikipedii, przetworzą dane, przygotowują środowisko do treningu modeli uczenia maszynowego oraz środowisko wdrożeniowe, wytrenują modele oraz zrealizują ich wdrożenie.

PRE-REQ: **Dobra znajomość podstaw Pythona, znajomość podstaw uczenia maszynowego z nadzorem oraz bez nadzoru**

DANE: **Wikipedia**

SERWISY: **Azure Functions, Azure Blob Storage**

GRUPA: **2 – 3 osoby**