

# Open Office dokumentumok elemzése

## Open Document Analyzer

A feladat egy Open Office dokumentumok elemzésére alkalmas modul kifejlesztése, amely beépül egy dokumentum management rendszerbe (DMS-be).

A termék segítségével dokumentumokhoz különböző információkat (metadata) csatolhatunk (pl. kulcsszavak, megjegyzések, könyvjelzők stb.), melyek megkönnyítik a keresést, és a későbbi feldolgozást. Lényeges funkció a kulcsszavak automatikus generálása, mely nagyban elősegíti a rendszerben tárolt dokumentumok egységes kezelését. A projekt Open Office dokumentumok elemzésére fókuszál, de később kiterjeszhető más formátumokra is.

Lehetne pontosítani ... milyen információk. hogyan könnyít.

Javaslom a projekt-be egy tesztelő modul beépítését is. azaz szükség van ..X.. (X>30) text dokumentumra ahhoz, hogy a funkcionalitást lehessen tesztelni

### A kezelt adatok

A modulnak Open Office dokumentumokat kell kezeljen, ezeknek a stílusadatait, a tartalmat, illetve a meta-adatait, vagyis a dokumentumokhoz hozzáfűzött pluszinformációkat. Ugyanakkor figyelembe kell vegye, hogy egy dokumentum több nyelven is íródhat.

### A felhasznált technológiák

A projekt kódja Java nyelvben íródik, Eclipse keretrendszert, framework -ot felhasználva.

### A projekt funkcionalitásai

#### 1. A elemzésre kerülő dokumentum beolvasása

- A bemeneti fájl helyes értelmezése
- A bemenet fájl megfelelő szétbontása
- Különálló, helyettesíthető szövegjellemző

#### 2. Többnyelvűség

felismeri a többnyelvű lehetőségek közül azt, amellyel operál?

- A dokumentum nyelve szerinti kulcsszógenerálás

- Kivételszavak, kivétel szótárak
- Angol, illetve magyar nyelv

### 3. Kulcsszavak generálása

- A generálandó kulcsszavak száma a rendszer paramétere
- A szavak előfordulás szerinti súlyozása
- A dokumentum formázásának figyelembevétele
- Nem kell felismernie azonos jelentéssel bíró szavakat, szinonimákat

### 4. A GUI

- A tesztelés, illetve a használathoz nyújt felületet
- Egyszerű
- Könnyen kezelhető

### 5. A kimenet

- A DMS egy belső formátuma
- XML alapú

### 6. Kiterjesztés más formátumokra

- A modulnak el kell fogadnia más standard dokumentum formátumokat is

## A projekt résztvevői

A projekten öt személy dolgozik.

Személyek	Munkakör
Pop Anita	fejlesztő
Fazakas Lóránd	fejlesztő
Lakatos Péter	fejlesztő
Pesti Pál	fejlesztő
Mihály Zsombor	fejlesztő, projekt menedzser