Komplex MI rendszerek házi feladat

# Problémafelvetés

A Steam nevű online játékplatform és webshop számos keresési opciót tartalmaz, lehet például szűrni címszavakra, kategóriákra és a felhasználók által megadott címkékre. Ezeket a címkéket a felhasználók szabadon adhatják meg játékokra, és lehetséges ezekre szavazni, azonban egyéb szemantikus információ nem tartozik hozzájuk, így például a „Horror” és a „Psychological horror” címkéknek semmi közük nincs egymáshoz. Ezt a hibát kívántam egy saját ontológia felhasználásával kiküszöbölni.

# Specifikáció

A projekt során egy java nyelvű programot készítek, amely letölti a Steam honlapjáról a legnépszerűbb címkéket, elkészít belőlük egy kezdetleges ontológiát, letölti Steam-ről a teljes játékgyűjteményt egy saját rdf adatbázisba, és képes ontológia alapján kiegészített keresést végezni az eltárolt játékok címkéin.

## Ontológia készítés

A programot a „CreateOntology” argumentummal jutunk az ontológia készítés funkcióhoz. Futás közben a <https://store.steampowered.com/tag/browse#global_492> url-t használom, ez az oldal a Steam legnépszerűbb felhasználói címkéit jeleníti meg a felhasználók számára. A letöltött oldalból kinyerem a listában szereplő címkéket, és ezek feliratai alapján készítek egyszerű OWL ontológiát, amely minden címkét egyedi osztályként tartalmaz. A megfelelő működéshez ezt az ontológiát kézzel kell kibővíteni.

Lehetőség van egy flag-et bebillenteni, hogy az ontológiához használt címkéket ne a weblapról, hanem egy csatolt .csv fájlból olvassa be. Ez a táblázat a <http://steamspy.com/tag/> honlapról származó 50 legtöbbet használt címkét tartalmazza.

## Adatbázis feltöltése

A programot a „FillDB” argumentummal jutunk az ontológia készítés funkcióhoz. A működés előfeltétele, hogy fusson egy Rdf4j szerver a localhost 8000-es portján. A teljes játék listát a (<http://api.steampowered.com/ISteamApps/GetAppList/v0001/>) Steam nemhivatalos API-járól kerül letöltésre. A lekérés eredmény egy Json objektum, ami minden Steam-en lévő artifact appId-jét tartalmazza. Ezek jelentős része nem elérhető, így azok nem is fognak bekerülni az adatbázisba. Az innen kapott appId listán végigiterálva a játékokhoz az adatok két helyről származnak, egyrészt az előbb említett nem publikus API-ról (<https://store.steampowered.com/api/appdetails?appids=8930>) le lehet kérni egy-egy appID-hez tartozó információkat, azonban itt nem szerepelnek a felhasználók által megadott címkék. A másik lekérés az appId-hez tartozó publikus steam oldalról történik (<https://store.steampowered.com/app/413150>), a letöltött html oldalból nyerem ki a címkéket.

Az adatbázisba {appId ?pred ?value} alakban kerülnek feltöltésre az adatok, ahol a ?pred a „title” „category” „tag” „type” és „metascore” értékeket veheti fel a ?value pedig a hozzá tartozó értékeket. Az adatbázisban minden IRI a <http://hu.komplexmi.e5q6ui#> névtér prefixszel van ellátva.