



Budapesti Műszaki és Gazdaságtudományi Egyetem  
Méréstechnika és Információs Rendszerek Tanszék

## Komplexesemény-feldolgozó (CEP) eszközök vizsgálata

**Rózsavölgyi Boton, IV. évf, (BSc) mérnök informatika szakos hallgató**

**Konzulens: Bergmann Gábor, MIT**

**Informatikai technológiák szakirány,**

**Rendszertervezés ágazat**

**Önálló laboratórium 2 összefoglaló**

**2016/17. II. félév**

A *complex event processing* (CEP), vagyis komplexesemény feldolgozás egy módszer, ami különböző (gyakran elosztott) rendszerekben beérkező atomi, időbélyeggel ellátott események folyamának az – általában valós idejű – feldolgozását, elemzését jelenti. Az eseményeket beérkezésük után logikai, térbeli és időbeli alapon összekapcsolhatjuk, belőlük komplex eseményekre következtethetünk. A CEP felhasználhatósága sokrétű; használható üzleti alkalmazásokban (pl. tőzsdei alkalmazások, hitelintézetek, kockázattertelő rendszerek, szállítmánykövetés), üzleti tevékenységet monitorozó eszközökhöz, nagy IT üzemeltető rendszerekhez (pl. komplex támadás felderítés, metrika kiértékelés), online visszaélések felderítéséhez (pl. fogadóirodák).

A szoftverpiacon sok különböző, CEP-re alkalmas eszköz áll rendelkezésre mind hagyományos, mind elosztott rendszerekre. A vizsgálatra az önálló laboratórium keretében a Drools-t választottam, mivel ez az eszköz egy java alapú, nyílt forráskódú szoftver, ami bármilyen java projektbe könnyen importálható, az Eclipse fejlesztőkörnyezetbe telepíthető plugin elérhető az Eclipse Marketplace-en. Natív, természetes nyelvre hasonlító, deklaratív szabályleíró nyelvvel rendelkezik, ami nem igazán emlékeztet az SQL-re, így kihívást jelentett, megismerése új tudást adott. A témát ezek a praktikus dolgok mellett érdekessége miatt is választottam, mivel szoftverfejlesztés ezen ága még viszonylag új, és potenciálisan sok érdekes problémára lehet vele megoldást találni.

Az önálló laboratórium keretében a Drools eszköz vizsgálatára a DEBS2015 Grand Challenge nevű feladat implementálását tűztem ki célul. A DEBS (ACM International Conference on Distributed Event-Based Systems) egy évente megrendezett nemzetközi konferencia elosztott eseményvezérelt rendszerek témakörében. Minden évben kiadnak egy – a konferencia témájával kapcsolatos – feladatot, amire csapatok nevezhetnek. A 2015-ös feladat egy taxitársasággal kapcsolatos, aminek olyan szoftverre van szüksége, ami valós idejű statisztikákat közöl az aktuális fuvarokról. A program olyan adatrekordokat dolgoz fel, amit a taxik küldenek be a központba egy fuvar végeztével. Ezek a rekordok tartalmaznak minden releváns információt; a taxi és a sofőr azonosítóját, az induló és a befejező koordinátákat és pontos időpontokat, a kifizetett fuvardíjat és borraalót, útdíjakat stb. A kiadott kihívás két részfeladatot definiált; az elsőben az elmúlt fél óra tíz legnépszerűbb útvonalát kellett kiszámolni és egy toplistán karbantartani, a másodikban pedig az elmúlt negyedóra legjövedelmezőbb környékeit hasonló módon, ahol a jövedelmezőség mérőszáma az elmúlt negyed órában az adott területen befejeződött fuvarok fuvardíjának és borraalójának összegének a mediánja elosztva az elmúlt fél órában a helyszínen befejeződött olyan fuvarok számával, ahol a sofőrnek még nem jelent meg a rendszerben újabb rekordja, vagyis a környéken fellelhető potenciálisan üres taxik száma. A területeket mindkét feladatban cellák reprezentálják (az elsőben 300\*300-as, a másodikban 600\*600-as felbontásban), amik New York térképének a releváns részének a felosztásával kaphatunk meg. A kihíváshoz egy 2 millió, -és egy 12 millió rekordot tartalmazó CSV fájlt adtak ki.

Először elkészítettem egy keretrendszert a tesztfájl beolvasására és feldolgozására, ami a megfelelő adatformátumokra konvertálja a különböző oszlopokat, illetve a megadott paraméterek alapján átszámolja a koordinátákat a kért cellákra. Ezt követte a szabályokat tartalmazó DRL fájlok első koncepciójának a kidolgozása, majd megvalósítása és kipróbálása. A szabályrendszer teszteléséhez JUnit tesztek hoztam létre, amit a DRL fájlok folyamatos fejlesztése és átdolgozása mellett szintén bővítettem, többször átalakítottam. Az elkészült szabályrendszereket többször lefuttattam a kiadott példafájlon, közben kisebb módosításokat végeztem, amíg a teljesítmény az elvártan megfelelően nem alakult. A mérések végül sikerrel zárultak, a második, bonyolultabb szabályrendszer is a valós idő harmada alatt dolgozta fel az eseményeket a legnagyobb terhelés alatt.