

# Véleménydetekció magyar nyelvű szövegek elemzése és gráfanalízis alapján

*Berend Gábor*

*Szegedi Tudományegyetem Természettudományi és Informatikai Kar*

*Konzulens: Farkas Richárd, Dr. Csirik János*

*Ph.D. hallgató, egyetemi tanár*

Napjainkban a fórumokon, blogokon található folyó szövegekből történő véleménykinyerés a természetes nyelvfeldolgozás egy igen nagy erővel kutatott szegmense, ahol az információ kinyeréssel ellentétben nem a dokumentumok tartalma képezi a vizsgálandó tárgyát, hanem a benne rejlő érzelmi töltet illetve vélemények detektálása. Számos alkalmazása képzelhető el egy véleménydetektáló rendszernek, úgy, mint fogyasztói elégedettség automatizált felmérése egy termékkel, termékcsoporttal kapcsolatban, vagy például politikai pártok, közszereplők támogatottságának megismerése. Mindehhez meg kell találni a szövegben rejlő úgynevezett pozitív-negatív polaritást, ami nemzetközi szinten is nyitott kérdés, hiszen a természetes nyelv nagyon sokféleképpen képes azt kifejezni.

Dolgozatomban egy, az emberi annotáció hatékonyságát megközelítő pontosságú véleménydetektáló rendszer kerül bemutatásra. A kutatási munka újszerűsége többek között abban rejlik, hogy ezt megelőzően még nem készült véleménykinyerő rendszer magyar nyelvű szövegekre. A vizsgált szövegek a magyaroszag.hu kormányzati portál fórumának a 2004. december 5-i kettős állampolgárságról szóló ügydöntő népszavazás témájában indított topikjának hozzászólásai voltak. A hozzászólásokat emberi erővel annotáltuk, a topikba hozzászólást író szerzők ezek alapján klasszifikálásra kerültek. A cél egy-egy hozzászóló polaritásának automatikus meghatározása volt.

A probléma megoldása folyamán különböző gépi tanulási módszereket alkalmaztam. A módszerek egy része a szövegből épített úgynevezett vektortér modellen alapul, ezen felül a predikció folyamán felhasználtam az egyes hozzászóláskora érkezett válaszreakciókból épített irányított gráfot is. A végső döntést egy szavazóalgoritmus hozza meg. Az így kapott eredmények a rendelkezésre álló viszonylagosan kis tanító adatbázis valamint a magyar nyelv különleges tulajdonságainak (szabad szórend, agglutináció) dacára is csupán néhány százalékkal maradtak el a fórum hozzászólásainak annotálása során tapasztalt annotátorok közötti egyetértési szinttől.