

MAGYAR GYAKORISÁGI SZÓTÁRAK A TARTALOMFELTÁRÁS SZOLGÁLATÁBAN

*Lengyelne Molnár Tünde, mtunde@ektf.hu
Eszterházy Károly Főiskola, Médiainformatika Tanszék*

Nagyon sok kutatás foglalkozik az információ kereső rendszerek hatékonyságának növelésével, az internetes keresőoldalak keresőmotorának fejlesztésével. Véleményem szerint azonban a tartalmi feltárás nemcsak az interneten megjelent információk áttekintését kell hogy jelentse, egy kutató-, oktató akkor is nehéz helyzetben van, ha a saját tudományterületén a nyomtatásban megjelent publikációkat szeretné figyelemmel kíséreni. Erre a problémára megoldást jelenthetnek a referátumok, melyek előállításának határt szab a készítő dokumentalisták véges kapacitása. Ezen helyzet javítását szolgálná saját kutatásom, mely egy magyar nyelvű offline kivonatoló program készítésére irányul.

A kivonat készítés automatizálásának lényegét a szöveg legfontosabb mondatainak meghatározásában látom, melynek előállítása a nyelvészet eredményeinek felhasználásával statisztikai alapon történik. A szignifikáns szavak listájának meghatározása során figyelembe kell venni az adott terület kifejezéseiből készített gyakorisági szótárakat. Ezen cél érdekében feltártam a Magyarországon eddig megjelent gyakorisági szótárakat. Előadásomban ezen szótárakat mutatom be részletesen, összehasonlítva gyakorisági adataikat, elérhetési módjaikat, illetve a korpusz, vagyis az alapul szolgáló szöveganyag nagyságát.