

AZ EMBERI KIVONATOLÁS STATISZTIKÁJA

Lengyelne Molnár Tünde
Eszterházy Károly Főiskola

Kulcsszavak: kivonatolás, szövegstatistika

Kutatási területem magyar nyelvű automatikusan kivonatoló program létrehozása, mely hiánytermék Magyarországon. Angol nyelvű szövegek esetén számos program létezik, melyek kvantitatív vagy kvalitatív elveket alkalmazva végzik el a lényeg kiemelését. Magyar nyelvű kivonatoló program létrehozásának alapjául *Krippendorff* tartalomelemző módszereit, *Luhn* elgondolásaira épülő szövegstatistikai iskola eredményeit, valamint több hazai nyelvészeti kutatás eredményeit is felhasználom. A program megírásához fontosnak tartom az emberi kivonatolás sajátosságainak vizsgálatát, úgy vélem, ez a terület eddig nem képezte kutatások alapjait.

Kutatásom célja feltárni az emberi kivonatolásban rejlő szabályszerűségeket, és az eredményeket felhasználni egy automatikusan kivonatoló program előállításához. Ehhez közel 350 fős mintán végeztem felmérést, melyben eltérő témájú cikkek leglényegesebb mondatainak kiválasztására kértem fel a résztvevőket, akik felsőoktatási intézmények hallgatóiból, illetve szakemberekből tevődtek össze.

Felmérésem során vizsgáltam, meghatározza-e a kivonatolást a készítő szakterülete, illetve a cikk témája, ezért a felmérés alapjául szakterületi és nem szakterületi cikkeket választottam. Elemeztem, mely mintacsoportok jelölik meg hasonlóan a lényeges mondatokat, összefügg-e a lényegkiválasztási technika a kivonatkészítő képzési szintjével.

Előadásomban szeretném bemutatni, milyen sajátosságok jellemzik az egyes mintacsoportok kivonatát, mely mintacsoportok kivonatánál mutatható ki összefüggés. Ezen túlmenően vizsgáltam, hasonlóan látják-e a szerzők a saját cikkük lényegét a cikket olvasó személyekkel. Előadásomban bemutatom, milyen viszonyban vannak a kapott kivonatok a szerzők saját maguk által megjelölt leglényegesebb mondataival.

A felmérés legfőbb eredménye, hogy megkísérel választ adni arra, hogy van-e kimutatható szabálykeresés a kivonatolás során.

A kutatás végső célja olyan program készítése, mely alkalmas lenne cikkek tartalmát tükröző mondatok kiválasztására. A program több területen használható, így például

- a nem referált szakterületi (pl. pedagógiai, könyvtartományi, nyelvészeti) adatbázisok cikkei absztraktjainak automatikus elkészítésére;
- az egyetemi/főiskolai könyvtárak szakdolgozatainak referált nyilvántartására;
- továbbá a jelentős digitalizálási projekteknek köszönhetően egyre több dokumentum lesz elérhető elektronikus formában, mely mindenképp maga után vonja az igényt egy kivonatoló programra.