

I. 312.

Fényes Balázs 10. o. t.

Budapest, Szerb A. Gimn.

feny.es.balazs@freemail.hu

A feladat vizsgálata és a talált eszközök

Az interneten a **tag cloud** (=címkefelhő) kifejezésre rákeresve próbáltam ilyen ábrák készítésére alkalmas eszközöket keresni. Ezeket *magyar* nyelvű szövegekkel teszteltem. A különböző oldalaknak meg lehetett adni egy nyers szöveget, amiből egy ábra vagy hasonló készült a szavakból úgy, hogy a különböző szavak a gyakoriságuknak megfelelően nagyobb mérettel rendelkeztek. Néhány oldal viszont szöveg helyett egy URL-címet kért, és a felhőben nem a weboldal szavai szerepeltek, hanem az oldal linkjei.

Általánosságban a következőket tapasztaltam:

- Mivel az általam talált eszközök angol nyelvűek voltak, ezért nem mindegyik hagyta a szavakon az ékezeteket, vagy más kódolásra alakították a szavakat, így a szavak olvashatatlanná váltak.
- A különböző viszonyszók (pl. az, be, nem) nyilvánvalóan a többi szóhoz képest túl sokszor szerepelnek. Ritka volt az olyan oldal, ahol ki lehetett választani a szöveg nyelvét, és még ritkább, amikor a magyart is. Viszont gyakran volt lehetőség eltávolítani egyes szavakat, de ez egy eléggé hosszadalmas folyamat lenne.
- Nagyrészt ezek a weboldalak nem egy ábrát állítottak elő, hanem egy formázott HTML szövegdobozt. Ami azért probléma, mert az ábrával ellentétben, ahol a programok megpróbálják a lehető legnagyobb mértékben kitölteni a területet, a szöveges felhőkben a sorok méretét a sor legnagyobb magasságú eleme határozta meg.
- Problémát jelent néhány oldalnál, hogy a címkefelhő megtekintéséhez Flash vagy Java szükséges, mivel ezek nincsenek meg minden hordozható eszközben (tabletek, okostelefonok).

Ezeket a címkefelhőt készítő weboldalakat látogattam meg:

- **TagCrowd** <http://tagcrowd.com/>
Egyáltalán nem támogatja az ékezetes karaktereket. Megadható a szöveg nyelve (tehát ezen nyelv gyakori szavait nem jeleníti meg); meglepő, hogy bár be lehet állítani magyar nyelvet, mégsem jelennek meg az ékezetek. Nem ábrát generál, hanem listát a gyakori szavakból (ez azért jó, mert képes a megfelelő szélességűre átméreteződni. Minden kék színű, ezt nem lehet változtatni. A szavak abc-sorrendben jelennek meg. Lehet PDF-be menteni.
- **Tagul** <http://tagul.com/>
Ábrákat hoz létre Flash alkalmazásként, ami beágyazható más weboldalakra, de akár PNG és SVG formátumban is elmenthető a kész mű. A használathoz regisztráció szükséges, ez arra jó, hogy a korábban létrehozott ábrákhoz később is hozzá lehet férni. Kezeli az ékezeteket, nem szűri a gyakori magyar szavakat, de gyakorlatilag minden testreszabható (még címkefelhő alakja is).
- **Wordle** <http://www.wordle.net/>
Ábrát hoz létre, magyar szavakat szűr, egyszerűen lehet szavakat eltávolítani. Weboldalra nem lehet az ábrát beágyazni, csak a linkjét, kell hozzá Java, a felhő alakját nem lehet testreszabni, emiatt mindig vízszintesen nyújtott alakú, sokszor egyes nagyméretű szavak kilógnak az ábrából, ezért sokszor kell frissíteni a felhőt ahhoz, hogy normálisan nézzen ki
- **TagCloud Generator** <http://www.tag-cloud.de/>

A szavakat egy 3D-s labdán ábrázolja, amit az egér mozgatásával lehet forgatni, nagyon kevés beállítási lehetőség, Flash alkalmazás, rosszul jeleníti meg az ékezetes karaktereket, nem szűr a gyakori magyar szavakra, nem lehet nyers szöveget bevinni, csak a weboldal címét, sokkal kevesebb szavat talált, mint amennyi valóban volt

- **TextTagCloud** <http://www.artviper.net/texttagcloud/>

Szöveges, magyar nyelven használhatatlan, kevés beállítási lehetőséggel rendelkezik, a szó színe jelképezi a gyakoriságát.

- **ToCloud** <http://www.tocloud.com/>

Szöveges, rossz karakterkódolással jeleníti meg a magyar szavakat, különböző szónak számítja azt a szót, aminek a végén idézőjel van, angol nyelvű szöveg esetén sem túl jó, mivel külön szónak veszi a szavak többes számát (pl. problem, problems – külön számolja), ha egy több szóból álló kifejezés többször szerepel, akkor nem a felhőben is egyben fog szerepelni a kifejezés, és nem vágja szét több szóra (pl. european grand prix -et egy szónak veszi).

- **WordSift** <http://www.wordsift.com/>

Szöveges, nem távolítja el a gyakori magyar szavakat, kevés szó, egyáltalán nincsenek beállítási lehetőségek, lassú.