

# Az *Idegen szavak magyarul* internetes szótár használatának elemzése

Kiss Gábor–Kiss Márton

Az elmúlt évtizedben sokat hallhattunk, olvashattunk arról, hogy a számítógép az élet oly sok területe mellett a szótárhasználatot is átalakítja, hiszen a számítógépes szótárakban nemcsak a címszavak ábécérendjében kereshetünk, hanem keresztbe-kasul. A jól elkészített keresőprogram lehetőséget biztosít arra, hogy a szócikkek belsejében lévő információkra is rákereshessünk. Kikereshetjük például egy értelmező szótár címszavai közül a minősítésekkel ellátott szavakat. A hétkötetes *Magyar nyelv értelmező szótárában* például néhány iparághoz köthető szakszavak száma a következő:

*asztalosipar* 57, *bányaipar* 164, *bőripar* 48, *cipészipar* 24, *malomipar* 15, *szövőipar* 200, *vegyipar* 448

A következő 21 szó mellett áll a „cukrászati” minősítés: *babliszt, bevonat, bonbon, csokoládémáz, desszert, fatörzs, fonálpróba, grillázs, habcsók, habtejszín, karamell, habtekercs, jég, kandíroz, marcipán, minyon, ostya, parfé, pille, praliné, sarokház*. Ezeket az adatokat az Arcanum Adatbázis Kft. által kiadott CD segítségével pillanatok alatt meghatározhattuk, míg a papír alapú szótár végigolvasása bizony hosszú időt vett volna igénybe.

Látjuk, hogy a számítógépes szótárhasználat számos előnnyel szolgál a felhasználó számára. Felvetődik a kérdés, hogy a szótárak szerkesztőinek is nyújt-e előnyt a gépi szótárhasználat, illetve a gépi szótárhasználat elemzése.

Füredi Ignác szótárszerző 1896-ban kiadott kis *Idegen szavak szótára* előszavában kissé hangzatosan így ócsárolja a korábbi szótárakat és dicséri saját gyűjteményét: *Régi panasz, hogy amit szótárainkban keresnek, az nincs meg bennük, és ami megvan, azt nem keresi senki*. Ezt az állítást több mint száz éven keresztül senki sem tudta sem cáfolni, sem igazolni. Azonban napjainkra, mikor már számítógépes és internetes szótárak is vannak, adódik a gondolat: a gép segítségével keressük meg a választ Füredi Ignác kijelentésére.

Vizsgálatunkhoz jó alap Tótfalusi István szerző *Idegen szavak magyarul* című szótára, amely papíralapú kiadvány formájában 2001-ben jelent meg, és azóta számos kiadást ért meg. A szótár 10 250 idegen szónak adja meg egy vagy több magyar megfele-

lőjét. (Ezen túl a szótár 3600 szónak a régies írásmódját is tartalmazza: pl. orthographia = ortográfia; ovatio = ováció; overall = overáll.) Igazából tehát magyarító szótár, hiszen nem hosszas definícióval, hanem magyar szavakkal értelmezi az idegen címszót, amint azt az alábbi részlet is mutatja:

**intakt:**

- ép, érintetlen, sértetlen
- szűz
- hiánytalan, csorbítatlan
- becsületes, feddhetetlen, megvesztegethetetlen

**intarzia:**

- faberakás, famozaik

**integer:**

- érintetlen, sértetlen
- feddhetetlen

**integráció:**

- egységesítés, egybefoglalás, összevonás

**integrál:**

- egységesít, egységbe foglal, összesít, összevon
- belefoglal

*Az Idegen szavak magyarul* című szótár a papíralapú szótár megjelenését követően négy évvel, 2005-ben felkerült a TINTA Könyvkiadó honlapjára ([www.tintakiado.hu](http://www.tintakiado.hu)) és bárhonnán használhatóvá vált a világhálón keresztül. Rövid időn belül az internetes változat igen népszerű lett, a magyarító szótárt sokan idegen szavak értelmező szótáraként kezdték el használni. Ma már számos ugrópont, link mutat rá. Természetesen ahhoz, hogy interneten keresztül hatékonyan lehessen használni a szótárt, meg kellett tervezni és el kellett készíteni a számítógépes keresőprogramot. Ennek az internetes szótárnak a használatát vizsgáltuk. Folyamatosan rögzítettük a keresés legfontosabb paramétereit, többek között azt is, hogy mely szavakat keresték.

A vizsgált időszak 164 napig tartott, 2007. október 1-jétől 2008. március 12-ig terjedt.

A kereső és adminisztrációs program négy mezőben feljegyezte minden egyes keresett szóról a következőket:

1. mező: a keresett szó benne van-e a szótárban vagy sem: 1 / 0
2. mező: a keresett szót a legördülő menüből kattintással választották ki, vagy beírták a rendelkezésre álló ablakba: b / k
3. maga a keresett szó
4. a keresés ideje: év, hó, nap, óra, perc formában

Részlet a paramétereket tartalmazó regisztrációs fájlból:

- ```
1 | k | dékán | 2008 01 13 11:52
1 | k | dekkol | 2008 01 13 11:52
0 | b | gorenje | 2008 01 13 11:54
1 | b | konklúzió | 2008 01 13 11:54
```

A vizsgált fenti 164 napban összesen 373 150 keresés történt a szótárban. Úgy vélem, ez elég nagy szám ahhoz, hogy következtetéseket vonjunk le a szótárhasználat mikéntjére. A közel négy százezer keresés azt jelenti, hogy naponta 2275, óránként 95, percenként pedig 1,6 idegen szó jelentését tudakolták ettől az elektronikus szótártól. A gépi feljegyzés szerint kattintással 190850 alkalommal, míg beírással 182300 esetben kerestek. Ha a 13850 címszót egyenletes érdeklődéssel keresték volna, akkor egy szót átlagosan 27 alkalommal néztek volna meg.

Természetes, hogy az egyes szavakat nem egyenlő arányban keresték. Éppen az az érdekes vizsgálatunkban, hogy a gépi keresés regisztrációjával megtudhatjuk a keresés eloszlását. Beírással 182300 alkalommal, míg kattintással 190850 esetben kerestek. Látható, hogy hozzávetőlegesen egyforma a két érték. A kattintásos keresés esetében véleményem szerint nem igazi tudatos keresésről van szó, hiszen ekkor csupán a legördülő menüben a gép által felkínált szavak közül lehet választani. A beírással keresés véleményem szerint már sokkal inkább igazi tudásszomjat takar.

Az elektronikus szótárban beírással kerestek, és megtalált leggyakoribb 25 szó listáját alább adjuk közre, és feltüntettük a keresések számát is:

*releváns 503; kognitív 401; szubjektív 305; paradigma 287; objektív 269; adekvát 254; kompetencia 242; koncepció 228; innováció 226; empirikus 222; interakció 217; identitás 216; aspektus 205; integráció 205; explicit 204; attitűd 193; szignifikáns 182; ambivalens 180; prioritás 177; kontextus 176; konvergencia 174; anomália 172; koherens 165; entitás 163; implicit 163.*

A szótár szerkesztője és közreadója számára igen hasznos az a lista, amelyben azok a szavak vannak felsorolva, melyeket kerestek a szótárhasználók, de a szótár nem tartalmazott. Az elektronikus szótárban beírással eredménytelenül kerestek leggyakoribb 25 szó listáját az alábbiakban olvashatjuk. Feltüntetve a sikertelen keresések számát is:

*deviáns 141, integrált 133, andragógia 114, diverzifikáció 112, interperszonális 109, desztináció 108, interdiszciplináris 105, holisztikus 104, sztereotípiá 102, deviancia 93, feminista 82, globalizáció 79, attitűd 79, adekvát 79, szubszidiaritás 78, celeb 75, status quo 74, szofisztikált 71, determinált 69, kompatibilis 62, kognitív 62, empatikus 62, implementáció 61, diszciplína 61, dereguláció 61, alternatíva 61, konnotáció 60.*

Rá kell mutatnunk, hogy az *attitűd* és a *kognitív* szavakat rossz helyesírással írták be a felhasználók, ezért nem ismerte fel a program a szót és nem adta meg jelentését. Továbbá az eredménytelen keresések egy jelentős része onnan is adódhat, hogy *Idegen szavak magyarul* szótár elsődleges funkciójánál fogva nem tartalmaz olyan idegen címszókat, melyeknek nincs egyetlen szavas magyar megfelelője, szinonimája.

A 182 300 beírással keresés 75 966 alkalommal volt sikeres (41%) és 106 335 alkalommal (59%) sikertelen. Meglepő a sikertelen keresések nagy száma, ez talán azt bizonyítja, hogy a szótár 13850 címszava kevés, a használóknak nagyobb szótárra van szükségük.

Tovább analizálva a fenti eredményt, megtudjuk, hogy a 75 966 sikeres keresés során a szótár 13 850 szavából csupán 6533 szót (47%) néztek meg a szótárban, és 7317 szótárbeli szót (53%) nem kerestek egyszer sem a 164 nap alatt, a szavak több mint fele feleslegesen van a szótárban!

A szótárban egyetlenegyszer sem keresett 25 db „b” betűvel kezdődő szó:

*background, backhand, backup, badminton, bagázs, bagett, balerina, balneoterápia, band, bandázs, banderole, banknóta, bankó, bankokrácia, bankrott, bantamsúly, baptisztérium, bar cod, barbarizmus, bárd, barkarola, barrel, bas-relief, bastonnád, basztard, bataillon.*

A sikertelen 106335 keresés során összesen 54113 különböző szónak szerették volna megtudni az értelmét. Ezek között szép számmal szerepelnek helyesírási hibával beírt szavak és olyanok is, melyek nem tartoznak a szótár profiljába. Ettől függetlenül meglepően nagy ez a szám, a statisztikai megoszlás a következő:

1117 szót több mint 10-szer kerestek hiába,

2144 szót 5–9 alkalommal kerestek hiába,

11 978 szót 2–4 alkalommal kerestek hiába,

38 874 szót egyetlenegyszer kerestek sikertelenül.

Természetesen kattintásnál minden szó benne van a gépi szótárban, hiszen ha nem lenne benne, akkor a gép fel sem kínálná. A 190850 kattintásos keresés sem egyenletesen oszlott meg a szótár címszavai között. Mivel a legördülő menü az „a” betűs szavaknál kezdődik, természetesen ezekre kattintás többször történt, mint más szavakra. Kihagyva az „ab” betűkapcsolattal kezdődő szavakat, az alábbi lista mutatja a kattintással legtöbbször keresett 25 címszót. Feltüntettük a kattintások számát is:

*integrál 234, diverzió 178, sztereotíp 176, individuális 175, integráció 153, kognitív 149, szofisztika 143, adekvát 142, determinál 139, integrálódik 133, szociális 133, adekvál 132, individualitás 131, konvenció 131, humánus 126, individuum 123, deviáció 114, bagatell 113, manifestál 113, szubszidiárius 112, implementál 109, manifestáció 109, transzcendens 105, individualizál 104, konstrukció 103.*

Elmondhatjuk, hogy a 373 150 keresésből összesen 266 816 volt sikeres (72%) és 106 334 sikertelen (28%), tehát a szótárban az eredményes keresések száma két és félszer több mint az eredménytelené.

Összefoglalva, táblázatos formában az alábbiakban szemléltetjük a szótárhasználat és a keresés adatait:

| AZ ÖSSZES KERESÉS SZÁMA |                 |                      |
|-------------------------|-----------------|----------------------|
| 373 150                 |                 |                      |
| KERESÉS BEÍRÁSSAL       |                 | KERESÉS KATTINTÁSSAL |
| 182 300 esetben         | 190 850 esetben |                      |
| EREDMÉNYES              | EREDMÉNYTELEN   | EREDMÉNYES           |
| 75 966 esetben          | 106 335 esetben | 190 850 esetben      |
| 6 533 szó               | 54 113 szó      | 13 850 szó           |

Több esetben is előfordult, hogy a felhasználó félreértette a szótár funkcióját. Mintha a gépet a használója naivan univerzális elektronikus agynak nézné. Néhány furcsa feltejt kérdés, azaz a beírás szövege az alábbiakban olvasható:

*Mit jelent az érelmeszesedés?*

*Mit nevezünk ereklyének?*

*Mel in ore, fel in core.*

*Mea culpa, mea maxima culpa.*

*majonézes krumpli*

*magas hangolású klarinét*

Tudjuk, hogy kényes a „csúnya” szavak szerepeltetése a szótárakban. A magyar szótáriródalom kissé álszemérmesen kezeli nyelvünknek ezeket az elemeit, annak ellenére, hogy a szótárhasználó gyakran elsőként ezeknél a szavaknál nyitja ki a szótárt. Még ebben az *Idegen szavak magyarul* című szótárban is voltak szép számmal „próbálkozók”. A *f...sz* szóval 36 alkalommal, a *p...a* szóval pedig 29 esetben „vizsgáztatták” a szótárt.

Egy erre a célra kifejlesztett kis programmal azt is feltérképezhetjük a regisztrációs file alapján, hogy egy-egy szónak a vizsgált időtartamban egyenletes-e a keresése, vagy vannak olyan napok, hogy egy-egy szót sokan keresnek. Az ún. ablaktologatós technikával végigpásztáztuk a regisztrált keresések sorát. Többek között négy időpontra figyeltünk fel, amelyben egy-egy szó keresésének a száma igen megugrott:

*casting*: 2008. január 9–12.

*in flagranti*: 2008. január 5.

*perzekútor*: 2008. január 28. 19 óra 25 perc.

*szegregáció*: 2007. december 9.

Ezek közül kettőre talán a valós magyarázatot is megtaláltuk. A *Casting* című élvezetes magyar filmet 2008. január 9-én mutatták be. A *perzekútor* szó 2008. január 28-án este feladványként szerepelt egy televíziós vetélkedőben. (A versenyző nem ismerte a szót, s a játék korai abbagyására kényszerült.) Arra, hogy a másik két szót miért keresték sokan a jelzett időpontban, nem találtuk még meg a választ. Mindenesetre meglepő és félelmetes, hogy a szótárhasználaton és a keresett szavakon keresztül is megtudhatunk információkat arról, hogy mi érdeklí embertársainkat egy adott időpontban.

A regisztrációs fájlunkból kiindulva felrajzolhatjuk, hogy a vizsgált 164 nap alatt a szótár használata, a feltett kérdések száma mennyire volt egyenletes. Az 1. számú ábra mutatja a használat dinamikáját. A vízszintes tengely az idő, a függőleges pedig a használatok számának a tengelye. Az ábra közepe felé jelentkező minimum hely 2007 karácsonyára esik.

Nemcsak azt nézhetjük meg, hogy napok szerint mi a keresések üteme, hanem arra is választ kaphatunk a regisztrációs fájlból kiindulva, hogy a nap 24 órájában milyen a keresések megoszlása. Ezt 2. a ábra szemlélteti. A vízszintes tengelyen a nap órái vannak feltüntetve, a függőlegesen pedig a keresések száma. Megfigyelhető, hogy 9–10 óra körül, mikor az emberek megérkeznek a munkahelyükre megugrik a keresések száma, ebédidőben 12–13 óra között pedig csökken. 18 óra felé az átmeneti csökkenés a hazautazás miatt van.

Összefoglalásként megállapíthatjuk Füredi Ignác 112 éves kijelentéséhez kapcsolódóan, hogy a gépi szótár felhasználói szokásainak vizsgálója három kérdésre pontos választ tud adni:

1. melyek azok a szavak a szótárban, melyeket sikeresen kerestek,
2. melyek azok a szavak a szótárban, melyeket nem kerestek,
3. melyek azok a szavak, melyek nincsenek benne a szótárban, de keresték.

Ezen túlmenően először elmondhatjuk, hogy a vizsgálat során külön kell választani azokat a kereséseket, melyeket a legördülő felkínált lehetőségek közül kattintással választanak ki a felhasználók azoktól, melyeket beírással tesznek fel a gépi szótárnak.

Másodszorra leszögezhetjük, hogy a szótárszerkesztő az internetes szótár kereséseinek regisztrálásával eddig soha nem látott információkhoz juthat hozzá. Folyamatosan javíthatja, bővítheti szótárát olyan szavakkal, melyeket valóban keresnek a felhasználók.

Harmadszorra elmondhatjuk, hogy a szótár közreadója a keresett szavakon keresztül szinte ördögi módon belelát az emberek fejébe, akár mindennapjaiba.

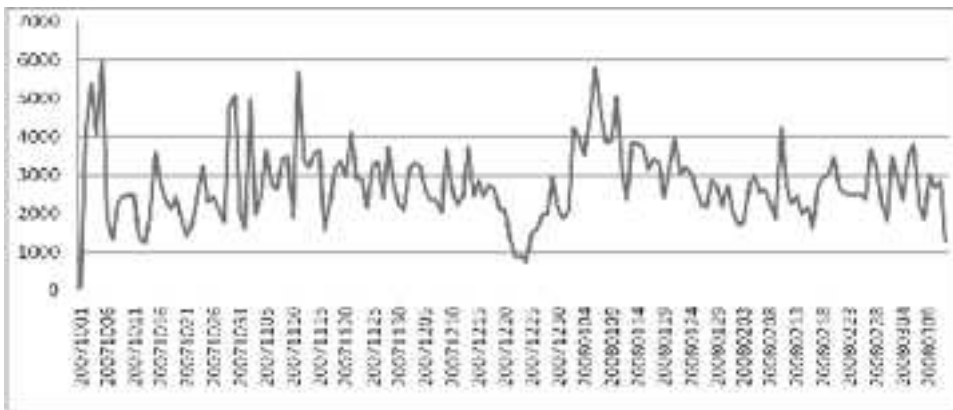
Negyedszerre levonhatjuk azt a tanulságot, hogy a felhasználók egy része félreérti a szótár funkcióját, és tévesen univerzális tudásbázisként próbálja használni az elektronikus szótárt és az azt működtető programot.

Sajnálatos módon a nagyszámú feltett kérdés és a rá adott válasz tudatformálásra és manipulációra is felhasználható. Ezért hallatlan nagy a szótárszerkesztő és közreadó felelőssége.

## Hivatkozások

- Füredi Ignác 1896. *Idegen szavak szótára*. Budapest  
Tótfalusi István 2001. *Idegen szavak magyarul*. Budapest: TINTA Könyvkiadó  
A magyar nyelv értelmező szótára. CD. Arcanum Adatbázis Kft., 2004.  
[http://www.tintakiado.hu/dictionary\\_idegenszotar.php](http://www.tintakiado.hu/dictionary_idegenszotar.php)

1. számú ábra



2. számú ábra

