Robot-lektor fejlesztése (részjelentés)

(Robot-lector – progress report)

Németh László (INNOREG), Pitlik László (MY-X team - 2014)

Kulcsszavak: objektivitás, makró, szövegbányászat, előfordulási gyakoriság, strukturáltság, hasonlóságelemzés, context free, benchmarking, ideálkeresés, versengő értékelés, nyelvtudomány, statisztika, mesterséges intelligenciák, szövegkorpusz

Kivonat: A robotlektor olyan szövegbányászati alapokra támaszkodó hasonlóságelemzési eljárás, mely képes hasznos (gyakorítandó) és káros (minimalizálandó) szóhasználati szokások aggregált értékelésére annak érdekében, hogy a tudományos szövegalkotást el lehessen választani a ködös mellébeszéléstől. A robotlektor a sakk-automatákhoz hasonlóan tények (szövegkorpuszok) és elméletek (szavak struktúraalkotásban betöltött támogató és akadályozó szerepe) optimális elegyeként került kialakításra.

Keywords: objectivity, macro, text mining, structuring, similarity analysis, context-free, benchmarking, competitive evaluation, statistics, artificial intelligence, text corpus, linguistic

Abstract: The robot-lector (based on similarity analyses and text mining logics) is capable of deriving statistics of the advantageous and disadvantageous words of a study and based on these descriptive data, the competitive comparison of papers can be executed in an automated way. The aim is to classify the-magic-of-words-objects and the scientific-valuable-ones. The robot-lector can be understood like a chess machine where facts (text corpus-elements) and rules (principles – e.g., role of words, linguistic structures) will lead to win (to a good study).

# Előzmények

* <https://miau.my-x.hu/miau/140/la140.doc>
* <http://miau.gau.hu/miau/162/FITt_v1.doc>
* <http://miau.gau.hu/miau/180/plagium_v2.docx>
* <http://miau.gau.hu/miau/181/etdk_2013_v4.doc>
* <http://miau.gau.hu/miau/183/etdk2013_v3.doc>

# Stratégiai célok

A TDK-dolgozatok minősítésén, rangsorolásán dolgozunk, annak érdekében, hogy a bírálói értékelés, a szubjektivitás mellett egy minél inkább objektívnek minősíthető, algoritmikus, gyors, számítógépes elemzéssel a dolgozat alapvető kritikai szempontjai értékelhetőek legyenek. Ezzel az objektivitásra törekvő módszerrel, a robot lektor, mely a dolgozat szóhasználata, szavak előfordulása, gyakorisága (a későbbiekben szomszédsági viszonyai), azok tudományossága és egyéb nyelvi attribútumok alapján tudja értékelni, hogy a TDK dolgozatok gondolatvezetésének strukturáltságát, azt, hogy a diákoknak mennyire sikerült kidolgozni, átadni formailag (a tartalmilag egyéb iránt nem vizsgált) üzenetet. A TDK-dolgozatok versengő minősítése során, ha a tudományosság megfelelő fokát eléri, akkor ideális esetben az objektivizáló értékelés az a szűrő lehet, mely segít feltárni, hogy egy második megmérettetés következhet-e, vagyis hogy a dolgozat részt vehet-e az pl. OTDK-n és/vagy érdemes-e emberi (szakértői) bírálatra…

# Operatív célok

Egy szoftver létrehozása a cél, mely segít pl. a TDK-dolgozatok elő-minősítésében, ill. a dolgozatírás során önkontrollként is szolgálhat. A tudományos értékkel bíró dolgozatok értékelésénél a bíráló szubjektivitás egy objektív mérőrendszerrel kerül kiegészítésre. Ez a lektor a dolgozat bírálók munkáját segítő alkalmazás, mely a szubjektív tartalmi értékelés előtt, mint egy előszűrőként vehet részt a TDK-dolgozat bírálatánál. A szoftver, mint cél kiválóan alkalmas IT-fókuszú képzések kapcsán önmagában is szakdolgozati/TDK-témának!

További technikai jellegű célkitűzés, hogy adott számú TDK-dolgozat értékelése minél kevesebb időt vegyen igénybe. Egy-egy dolgozat értékelésére a bírálók elvileg komoly időt fordítanak. A bírálók munkájának könnyítése során a robot–lektor rövid idő alatt a dolgozatról egy áttekintő képet ad, statisztikai összegzés alapján közelítő jellegű információt például arról, hány táblázatot, ábrát tartalmaz a tanulmány (ill. ezek a kifejezések hányszor kerültek említésre), hányszor szerepel pl. az elemzés, művelet, összehasonlítás, vizsgálat, stb. kifejezéstömeg, ami a munka minőségére/strukturáltságára utal. A hatékonyságnövelés fentebb leírt logikája megfelel a KNUTH-i elvnek: vagyis tudás az, ami forráskódba átírható!

TDK-dolgozatokkal szemben kiemelt elvárás a célul választott probléma tudományos igényű megközelítése és ennek megfelelő dokumentáltsága. Hiába egy-egy innovatív TDK dolgozat, mely új konstrukciós, tervezési, technológiai vagy vizsgálati eljárást, illetve annak egy részét dolgozott ki, amelyek a gyakorlatban hasznosíthatók, ha ennek bemutatása ködös, kontúrtalan. A jelentős társadalmi vagy ipari hasznosság esetleg azért nem hasznosul, mert a szerző elhitelteleníti saját munkáját saját stílusjegyein keresztül.

TDK-dolgozatok mellett természetesen bármilyen más írásmű értékelése is hasonló módon lehetséges!

# A megoldás alapjai

Szövegbányászati módszerek alkalmazása révén lehet a végeredményt megkapni, az emberi tényező minimalizálásával/kizárásával. Egy olyan megoldást kerül a bírálók részére biztosításra, mely a dolgozat áttekintéséhez egy előzetes szűrést biztosít, a dolgozatban használt kifejezések, szavak tudományosságra utaló használata révén, majd ezen statisztikák alapján történik a dolgozatok versengő értékelése, a norma alattiság, normaszerűség és norma-felettiség becslése.

Az eddig szereplő kulcsszavakat a szerzők korábbi elméleti (nyelvészeti) és gyakorlati (a dolgozatok online elérhető szövegkorpuszára támaszkodó tapasztalatai határozták meg, de ezek folyamatosan bővülő kifejezés-halmazok. Cél az összes logikai minősítéssel bírni képes szó beillesztése a robot-lektorba. Minden kifejezés egy-egy dolgozatra/kivonatra jellemző találati mennyisége rögzítésre kerül egy adatbázisban. Jelenleg a potenciális kifejezések gyűjtése, mérlegelése folyik, pl. melyik bír kellő logikai tartalommal a dolgozat minősítésére, illetve, hogy ezen szavakból a minél több, vagy a minél kevesebb a célravezető az ideál leírása szempontjából.

Az adatbázisunkat és az algoritmust magát a kísérleti fázisban egyelőre xls-alapon hoztuk létre, ahol az attribútumok a keresésben részvevő karaktersorok/kulcsszavak/szócsoportok. A sorfejlécbe a vizsgált dolgozat URL-je, kifejezés iránya, kifejezés csoportja és a kifejezés alkotja.

Jelenleg 492 attribútumunk van (1. számú melléklet), s ezeket 18 csoportra bontottuk szét logikai jelentésük/üzenetük szerint. Mindegyik csoportnak adtunk egy irányultságot és indoklást: T vagy K. Minél nagyobb találati szám a mérvadó, mint például a "logikai következtetésre utalás" című csoportban, akkor T, ha a minél kisebb találati szám a célravezető, mint például a "bizonytalanságra utalás" című csoportban K jelzés kerül mellé. Minden kifejezéshez tartozik egy irányultság és egy csoport, úgyhogy a kifejezés és a csoport irányultság nem lehet különböző. A vizsgált dokumentumot belehelyeztük az xls-be oly módon, hogy minden cella egyetlen szót tartalmaz a dokumentumból, és így lehetőség nyílt a keresési macro alkalmazására az xls-ben.

# Alternatív megoldások

Először hagyományos Ctrl+F keresés segítségével töltöttük ki az adatbázisunkat, ami túl sok időt vett igénybe 47’33” = 2853, amit stopperral mértünk. A kivitelezés hatékonyságát makrók segítségével csökkentettük 8” –re, azaz a 365-öd részére, vagyis 2 nagyságrenddel felgyorsult a folyamat egyetlen egy sms-nyi makró hatására – az elvárások szerint….



1. Ábra: Alapadatok (OAM) – forrás: saját ábrázolás

Az 1. ábra egyes quasi véletlenszerűen kiválasztott TDK-dolgozatok leíró statisztikáinak sorszám-nézetét mutatja be, ahol az objektumok (sorok) további dimenziókkal (pl. évjárat, reál/humán-státusz, ill. MYX-kötődés) vannak ellátva.



1. Ábra: Relatív adatok – forrás: saját ábrázolás

A 2. ábra a 3. ábra nyersadatainak 1000 karakterre jutó, azaz a mindenkori mű hosszával relativált értékeit tartalmazza. A 2. ábra egyben az 1. ábra sorszámozásának alapja.



1. Ábra: A szövegbányászati alapon feltárt nyersadatok – forrás: saját ábrázolás

# Eredmények

Az anti-diszkriminatív hasonlóságelemzés az 1. ábra alapján két modell-rétegben jut el az egyes objektumok idealitás-indexének becsléséhez, mert az első körben a COCO Y0 eljárás feltárt egy olyan attribútum-részhalmazt, mely alapján minden objektum lehetett másként egyforma értékesnek tekinthető (vö. <https://miau.my-x.hu/myx-free/>).

A második elemzési rétegben már csak 3 darab attribútum maradt (vö. sárga jelzés a 4. ábrán).



1. Ábra: A kétrétegű modell rétegei – forrás: saját ábrázolás
2. Ábra: Az idealitás-indexek időbeli alakulása

Az 5. ábra felső részén a teljes objektum-halmaz idealitás-trendje látható, ahol a trendfüggvény meredeksége lényegében nulla, azaz nincs javulásra utaló jel.

A MYX-kötődésű objektum esetén (vagyis az összes objektum egyik részhalmazára vonatkozóan) a trend masszívan pozitív, vagyis a szerzők által befolyásolt alkotási folyamat minőségre gyakorolt hatása triviálisan felismerhető.



1. Ábra: Az idealitás-indexek objektumonként – forrás: saját ábrázolás

A 6. ábra alapján az összes objektum leíró adatai és idealitás-indexei értelmezhetők. Látható, hogy az 1000024.1 pontos idealitás-maximum nem MYX-s objektum kapcsán állt elő, tehát lehet jó dolgozatot írni, csak akarni kell.

 

1. Ábra: A myx-kötődés és a humán/reál-vízválasztó hatása az idealitásra – forrás: saját ábrázolás

A 7. ábra alapján vélelmezhető, hogy reál területen a szómágia negatív hatásai kisebbek, s a MYX-kötődés is természetesen erősíti a minőséget. A MYX-kötődés azonban nem csak reál témakörökhöz kötődik.

# Konklúziók

Ha a mindösszesen halmaz eredő-trendje 0, s az egyik részhalmazának trendje pozitív, akkor a fennmaradó halmaz, vagyis a nem MYX-kötődésű objektumok kapcsán a trend masszívan negatív kell, hogy legyen, ami azt jelenti, hogy azok az alkotói közösségek, ahol az írásművek minőségmenedzsmentje nem kellően fókuszált, nem kellően tudatos, ott az idő múlásával a minőség önkatalitikus módon csökken/erodálódik , hiszen az elődök által sem túl magasra tett színvonalból „okulva” a mindenkori utódok apróbb nagyobb további hiányosságok még elfogadhatónak ítélésével negatív, dekadens spirált indítanak be.

A versengő objektum-értékelés, az alkotások evolúciója tehát nem következik be magától, ezért tenni kell. Tenni pedig az értékelés automatizálásán és objektivizálásán keresztül lehet a legtöbbet és a leghatékonyabban!

További részletek: <https://miau.my-x.hu/miau/270/robotlektor_oam_v2.xlsx>

# Mellékletek

1. számú melléklet: a kulcsszavak aktuális halmaza

|  |  |  |
| --- | --- | --- |
| Kifejezés | Irányultság | Csoport |
| mert | T | A |
| mivelhogy | T | A |
| hiszen | T | A |
| ugyanis | T | A |
| mivel | T | A |
| merthogy | T | A |
| de | T | A |
| következésképpen | T | A |
| szóval | T | A |
| különösképpen | T | A |
| különösképp | T | A |
| különösen | T | A |
| módfelett | T | A |
| tehát | T | A |
| megfejtés | T | A |
| eléggé | T | A |
| így | T | A |
| így hát | T | A |
| kifejezetten | T | A |
| egyszóval | T | A |
| más szóval | T | A |
| mármint | T | A |
| úgymint | T | A |
| mégpedig | T | A |
| éspedig | T | A |
| nevezetesen | T | A |
| azaz | T | A |
| vagyis | T | A |
| úgyhogy | T | A |
| voltaképpen | T | A |
| magyarán | T | A |
| nos | T | A |
| ergo | T | A |
| következésképp | T | A |
| következtetés | T | A |
| konklúzió | T | A |
| tanulság | T | A |
| konzekvencia | T | A |
| következően | T | A |
| bizonyít | T | A |
| igazol | T | A |
| beigazol | T | A |
| alátámaszt | T | A |
| bebizonyít | T | A |
| indok | T | A |
| ürügy | T | A |
| ok | T | A |
| indoklás | T | A |
| magyarázat | T | A |
| bizonyíték | T | A |
| argumentum | T | A |
| érv | T | A |
| alap | T | A |
| csomó | K | A1 |
| fáma | K | A1 |
| rahedli | K | A1 |
| milliányi | K | A1 |
| irdatlan mennyiségű | K | A1 |
| jó néhány | K | A1 |
| élmény | K | A1 |
| seregnyi | K | A1 |
| borzalmasan | K | A1 |
| töméntelen | K | A1 |
| nagy mennyiségű | K | A1 |
| millió | K | A1 |
| megszámlálhatatlan | K | A1 |
| fikarcnyi | K | A1 |
| csepp | K | A1 |
| csöpp | K | A1 |
| átkozottul | K | A1 |
| parányi | K | A1 |
| nyúlfarknyi | K | A1 |
| csipet | K | A1 |
| csipetnyi | K | A1 |
| leheletnyi | K | A1 |
| szemernyi | K | A1 |
| hangyányi | K | A1 |
| ici-pici | K | A1 |
| pirinyó | K | A1 |
| őrülten | K | A1 |
| merőben | K | A1 |
| veszettül | K | A1 |
| falatnyi | K | A1 |
| zömében | K | A1 |
| nagy általánosságban | K | A1 |
| erőteljesen | K | A1 |
| rohamosan | K | A1 |
| zömmel | K | A1 |
| nagyobbrészt | K | A1 |
| fantasztikusan | K | A1 |
| szörnyen | K | A1 |
| jórészben | K | A1 |
| jórészt | K | A1 |
| javarészben | K | A1 |
| túlnyomóan | K | A1 |
| többségében | K | A1 |
| többnyire | K | A1 |
| befolyás | K | A1 |
| igencsak | K | A1 |
| nagyrészt | K | A1 |
| rendszerint | K | A1 |
| általában | K | A1 |
| legtöbbször | K | A1 |
| jobbára | K | A1 |
| túlnyomórészt | K | A1 |
| leginkább | K | A1 |
| leggyakrabban | K | A1 |
| legfőképp | K | A1 |
| főképp | K | A1 |
| sok esetben | K | A1 |
| számos esetben | K | A1 |
| gyakorta | K | A1 |
| néha | K | A1 |
| olykor-olykor | K | A1 |
| egyszer-egyszer | K | A1 |
| alkalomadtán | K | A1 |
| elvétve | K | A1 |
| néhanapján | K | A1 |
| akkoriban | K | A1 |
| valamikor | K | A1 |
| valaha | K | A1 |
| hajdan | K | A1 |
| annak idején | K | A1 |
| egykor | K | A1 |
| hajdanán | K | A1 |
| réges rég | K | A1 |
| akkortájt | K | A1 |
| egykoron | K | A1 |
| hébe-hóba | K | A1 |
| időközönként | K | A1 |
| talán | K | A1 |
| tán | K | A1 |
| netán | K | A1 |
| esetlegesen | K | A1 |
| esetleg | K | A1 |
| netalántán | K | A1 |
| meglehet | K | A1 |
| éppenséggel | K | A1 |
| valószínűleg | K | A1 |
| saccolás | K | A1 |
| roppant módon | K | A1 |
| roppantul | K | A1 |
| roppant | K | A1 |
| borzasztóan | K | A1 |
| cefetül | K | A1 |
| veszettül | K | A1 |
| brutálisan | K | A1 |
| elképesztően | K | A1 |
| eszméletlenül | K | A1 |
| jócskán | K | A1 |
| roppant mód | K | A1 |
| irdatlanul | K | A1 |
| piszkosul | K | A1 |
| rettenetesen | K | A1 |
| irtózatosan | K | A1 |
| kegyetlenül | K | A1 |
| számtalanszor | K | A1 |
| félig-meddig | K | A1 |
| közelmúltban | K | A1 |
| többé-kevésbé | K | A1 |
| iszonyatosan | K | A1 |
| pláne | K | A1 |
| már-már | K | A1 |
| maroknyi | K | A1 |
| jóformán | K | A1 |
| kvázi | K | A1 |
| durván számítva | K | A1 |
| úgy-ahogy | K | A1 |
| tipp | K | A1 |
| keresés | T | A2 |
| konklúzió | T | A2 |
| kutatás | T | A2 |
| tanulság | T | A2 |
| igazol | T | A2 |
| ergo | T | A2 |
| alátámaszt | T | A2 |
| elképzelés | T | A2 |
| bizonyít | T | A2 |
| információ | T | A2 |
| eredmény | T | A2 |
| ábra | T | A3 |
| adat | T | A3 |
| adatállomány | T | A3 |
| adatbank | T | A3 |
| adatbázis | T | A3 |
| adatgyűjtemény | T | A3 |
| tény | T | A3 |
| adatelemzés | T | A3 |
| függvény | T | A3 |
| modell | T | A3 |
| információ | T | A3 |
| illusztráció | T | A3 |
| rajz | T | A3 |
| alátámaszt | T | A3 |
| következmény | T | A3 |
| bizonyít | T | A3 |
| kutatás | T | A3 |
| definiálás | T | A3 |
| definíció | T | A3 |
| elmélet | T | A3 |
| diagram | T | A3 |
| direkt | T | A3 |
| elv | T | A3 |
| eredmény | T | A3 |
| meghatározás | T | A3 |
| előterjesztés | T | A3 |
| megfejtés | T | A3 |
| érv | T | A3 |
| fogalom | T | A3 |
| tapasztalat | T | A3 |
| konklúzió | T | A3 |
| konzekvencia | T | A3 |
| táblázat | T | A3 |
| tudnivaló | T | A3 |
| összeg | T | A3 |
| útmutatás | T | A3 |
| végeredmény | T | A3 |
| végösszeg | T | A3 |
| végkifejlet | T | A3 |
| **.** | T | A4 |
| , | T | A4 |
| ; | T | A4 |
| ! | T | A4 |
| : | T | A4 |
| / | T | A4 |
| - | T | A4 |
| + | T | A5 |
| ^ | T | A5 |
| ÷ | T | A5 |
| × | T | A5 |
| - | T | A5 |
| % | T | A5 |
| 0 | T | A6 |
| 1 | T | A6 |
| 2 | T | A6 |
| 3 | T | A6 |
| 4 | T | A6 |
| 5 | T | A6 |
| 6 | T | A6 |
| 7 | T | A6 |
| 8 | T | A6 |
| 9 | T | A6 |
| I | T | A6 |
| II | T | A6 |
| III | T | A6 |
| IV | T | A6 |
| V | T | A6 |
| VI | T | A6 |
| VII | T | A6 |
| VIII | T | A6 |
| IX | T | A6 |
| X | T | A6 |
| XV | T | A6 |
| XVI | T | A6 |
| XVII | T | A6 |
| XVIII | T | A6 |
| XIX | T | A6 |
| XX | T | A6 |
| XXI | T | A6 |
| L | T | A6 |
| C | T | A6 |
| D | T | A6 |
| M | T | A6 |
| definíció | T | A7 |
| definiálás | T | A7 |
| argumentum | T | A7 |
| hierarchia | T | A7 |
| hipotézis | T | A7 |
| metamorfózis | T | A7 |
| impresszió | T | A7 |
| impulzus | T | A7 |
| input | T | A7 |
| output | T | A7 |
| konklúzió | T | A7 |
| konzekvencia | T | A7 |
| kritikus | T | A7 |
| kvázi | T | A7 |
| procedúra | T | A7 |
| szimuláció | T | A7 |
| transzformáció | T | A7 |
| diagram | T | A8 |
| ábra | T | A8 |
| kimutatás | T | A8 |
| rajz | T | A8 |
| táblázat | T | A8 |
| fotográfia | T | A8 |
| fotó | T | A8 |
| egyenesen | T | A8 |
| makett | T | A8 |
| skála | T | A8 |
| kép | T | A8 |
| besorolás | T | A8 |
| sorszám | T | A8 |
| riport | T | A8 |
| statisztika | T | A8 |
| grafika | T | A8 |
| grafikon | T | A8 |
| művelet | T | A8 |
| illusztráció | T | A8 |
| modell | T | A8 |
| sablon | T | A8 |
| szimuláció | T | A8 |
| bizonyít | T | A9 |
| bizonyíték | T | A9 |
| célcsoport | T | A9 |
| célközönség | T | A9 |
| célszerű | T | A9 |
| célzatosan | T | A9 |
| döntően | T | A9 |
| elképzelés | T | A9 |
| előnyös | T | A9 |
| foganat | T | A9 |
| előreláthatatlan | T | A9 |
| készakarva | T | A9 |
| jól | T | A9 |
| eredmény | T | A9 |
| érték | T | A9 |
| felhasználható | T | A9 |
| százalék | T | A9 |
| információ | T | A9 |
| szemlélet | T | A9 |
| kutatás | T | A9 |
| tanács | T | A9 |
| adatbázis | T | A9 |
| adatállomány | T | A9 |
| alaposan | T | A9 |
| benyomás | T | A9 |
| megfejtés | T | A9 |
| megoldás | T | A9 |
| vélemény | T | A9 |
| kimutatás | T | A9 |
| keresés | T | A9 |
| kimenetel | T | A9 |
| ajánlat | T | A9 |
| termés | T | A9 |
| statisztika | T | A9 |
| nonprofit | T | A9 |
| hatás | T | A9 |
| következmény | T | A9 |
| közérdekű | T | A9 |
| közhasznú | T | A9 |
| nyereség | T | A9 |
| haszon | T | A9 |
| profit | T | A9 |
| mindent összevetve | T | B |
| tehát | T | B |
| így | T | B |
| kimutatás | T | B |
| kimenetel | T | B |
| így hát | T | B |
| egyszóval | T | B |
| más szóval | T | B |
| mármint | T | B |
| úgymint | T | B |
| alaposan | T | B |
| mégpedig | T | B |
| ajánlat | T | B |
| indítvány | T | B |
| értelmezés | T | B |
| értelmében | T | B |
| értesítés | T | B |
| túlságosan | T | B |
| igazol | T | B |
| kezdeményezés | T | B |
| kiváltképpen | T | B |
| kiváltképp | T | B |
| ráhatás | T | B |
| behatás | T | B |
| fejlemény | T | B |
| magyarán | T | B |
| ajánlás | T | B |
| beajánlás | T | B |
| javallat | T | B |
| javaslat | T | B |
| reakció | T | B |
| folyomány | T | B |
| kimondottan | T | B |
| válasz | T | B |
| felvetés | T | B |
| meggyőződés | T | B |
| sorrend | T | B |
| szándékosan | T | B |
| jelentősen | T | B |
| számszerűsítés | T | B |
| kiugróan | T | B |
| tudatosan | T | B |
| magyarázat | T | B |
| megoldás | T | B |
| mert | T | B |
| hatás | T | B |
| mivelhogy | T | B |
| kihatás | T | B |
| hiszen | T | B |
| ugyanis | T | B |
| mivel | T | B |
| merthogy | T | B |
| felelet | T | B |
| rangsor | T | B |
| következmény | T | B |
| következésképpen | T | B |
| szóval | T | B |
|  sok | K | B1 |
| tömérdek | K | B1 |
| számtalan | K | B1 |
| rengeteg | K | B1 |
| ezernyi | K | B1 |
| számos | K | B1 |
| megannyi | K | B1 |
| temérdek | K | B1 |
| számottevő | K | B1 |
| nagyon sok | K | B1 |
| csomó | K | B1 |
| rahedli | K | B1 |
| milliányi | K | B1 |
| irdatlan mennyiségű | K | B1 |
| jó néhány | K | B1 |
| végtelen | K | B1 |
| töméntelen | K | B1 |
| nagyszámú | K | B1 |
| seregnyi | K | B1 |
| nagy mennyiségű | K | B1 |
| millió | K | B1 |
| megszámlálhatatlan | K | B1 |
| kevés | K | B1 |
| kicsi | K | B1 |
| minimális | K | B1 |
| kis mértékű | K | B1 |
| elenyésző | K | B1 |
| fikarcnyi | K | B1 |
| csepp | K | B1 |
| csöpp | K | B1 |
| némi | K | B1 |
| maroknyi | K | B1 |
| mérhetetlenül | K | B1 |
| parányi | K | B1 |
| kis | K | B1 |
| nyúlfarknyi | K | B1 |
| csipet | K | B1 |
| csipetnyi | K | B1 |
| pici | K | B1 |
| árnyalatnyi | K | B1 |
| távol | K | B1 |
| messze | K | B1 |
| apró | K | B1 |
| leheletnyi | K | B1 |
| szemernyi | K | B1 |
| hangyányi | K | B1 |
| részben | K | B1 |
| ici-pici | K | B1 |
| pirinyó | K | B1 |
| iciri-piciri | K | B1 |
| falatnyi | K | B1 |
| több | K | B1 |
| kevese | K | B1 |
| zömmel | K | B1 |
| nagyobbrészt | K | B1 |
| jórészben | K | B1 |
| jórészt | K | B1 |
| javarészben | K | B1 |
| túlnyomóan | K | B1 |
| többségében | K | B1 |
| zömében | K | B1 |
| nagyrészt | K | B1 |
| általában | K | B1 |
| legtöbbször | K | B1 |
| jobbára | K | B1 |
| nagy általánosságban | K | B1 |
| túlnyomórészt | K | B1 |
| leggyakrabban | K | B1 |
| főképp | K | B1 |
| többnyire | K | B1 |
| gyakran | K | B1 |
| sokszor | K | B1 |
| rendszerint | K | B1 |
| számtalanszor | K | B1 |
| többször | K | B1 |
| sűrűn | K | B1 |
| sok esetben | K | B1 |
| rendszeresen | K | B1 |
| több esetben | K | B1 |
| megannyiszor | K | B1 |
| gyakorta | K | B1 |
| néha | K | B1 |
| olykor-olykor | K | B1 |
| egyszer-egyszer | K | B1 |
| elvétve | K | B1 |
| közel | K | B1 |
| olykor | K | B1 |
|  egyszer  | K | B1 |
| egy ízben | K | B1 |
| esetenként | K | B1 |
| kevesebb | K | B1 |
| egyszer-egyszer | K | B1 |
| ritkán | K | B1 |
| csekély | K | B1 |
| alkalomadtán | K | B1 |
| hébe-hóba | K | B1 |
| néhanapján | K | B1 |
| időközönként | K | B1 |
| hasznos | T | B2 |
| megfelelő | T | B2 |
| hasznot hajtó | T | B2 |
| előnyös | T | B2 |
| hasznavehető | T | B2 |
| hasznosítható | T | B2 |
| felhasználható | T | B2 |
| alkalmazható | T | B2 |
| praktikus | T | B2 |
| célszerű | T | B2 |
| használható | T | B2 |
| ellenben | T | B3 |
| viszont | T | B3 |
| csakhogy | T | B3 |
| ám | T | B3 |
| ámde | T | B3 |
| azonban | T | B3 |
| mégis | T | B3 |
| de | T | B3 |
| ennek ellenére | T | B3 |
| bár | T | B3 |
| habár | T | B3 |
| pedig | T | B3 |
| ugyan | T | B3 |
| döntően | T | B4 |
| jellemzően | T | B4 |
| főként | T | B4 |
| főleg | T | B4 |
| kiváltképp | T | B4 |
| gyakran | T | B4 |
| sokszor | T | B4 |
| számtalanszor | T | B4 |
| többször | T | B4 |
| hetente | T | B4 |
| naponta | T | B4 |
| évenként | T | B4 |
| óránként | T | B4 |
| megannyiszor | T | B4 |
| számos esetben | T | B4 |
| gyakorta | T | B4 |
| néha | T | B4 |
| olykor-olykor | T | B4 |
| egyszer-egyszer | T | B4 |
| alkalomadtán | T | B4 |
| hébe-hóba | T | B4 |
| elvétve | T | B4 |
| olykor | T | B4 |
| néhanapján | T | B4 |
| esetenként | T | B4 |
| korábban | T | B4 |
| régebben | T | B4 |
|  egyszer  | T | B4 |
| nemrég | T | B4 |
| múltkor | T | B4 |
| ezelőtt | T | B4 |
| egy ízben | T | B4 |
| régen | T | B4 |
| anno | T | B4 |
| akkoriban | T | B4 |
| hajdanában | T | B4 |
| valaha | T | B4 |
| hajdan | T | B4 |
| annak idején | T | B4 |
| egykor | T | B4 |
| hajdanán | T | B4 |
| réges rég | T | B4 |
| akkortájt | T | B4 |
| egykoron | T | B4 |
| modell | T | B5 |
| mintadarab | T | B5 |
| minta | T | B5 |
| makett | T | B5 |
| mintapéldány | T | B5 |
| sablon | T | B5 |
| séma | T | B5 |
| szimuláció | T | B5 |
| alapján | T | B5 |
| besorolás | T | B5 |
| ranglista | T | B5 |
| szerint | T | B5 |
| kutatás | T | B5 |
| mintájára | T | B5 |
| akarattal | T | B5 |
| szántszándékkal | T | B5 |
| eljárás | T | B5 |
| átváltozás | T | B5 |
| átalakulás | T | B5 |
| felvetés | T | B5 |
| procedúra | T | B5 |
| igazán | T | B5 |
| közvetlenül | T | B5 |
| ügymenet | T | B5 |
| folyamat | T | B5 |
| művelet | T | B5 |
| menet | T | B5 |
| ábra | T | B5 |
| kép | T | B5 |
| ötlet | T | B5 |
| grafika | T | B5 |
| illusztráció | T | B5 |
| álláspont | T | B5 |
| nézet | T | B5 |
| állásfoglalás | T | B5 |
| grafikon | T | B5 |
| felvétel | T | B5 |
| rajz | T | B5 |
| fénykép | T | B5 |
| fotó | T | B5 |
| fotográfia | T | B5 |
| diagram | T | B5 |
| eljárás | T | B5 |
| eredmény | T | B5 |
| következmény | T | B5 |
| fejlemény | T | B5 |
| megoldás | T | B5 |
| értelmében | T | B5 |
| végkifejlet | T | B5 |
| értelmezés | T | B5 |
| felhasználható | T | B5 |
| feltételezés | T | B5 |
| szemlélet | T | B5 |
| foganat | T | B5 |
| skála | T | B5 |
| inger | T | B5 |
| adat | T | B5 |
| hír | T | B5 |
| válasz | T | B5 |
| infó | T | B5 |
| túlzottan | T | B5 |
| meglátás | T | B5 |
| tájékoztatás | T | B5 |
| értesítés | T | B5 |
| fáma | T | B5 |
| információ | T | B5 |
| ismertető | T | B5 |
| tudnivaló | T | B5 |
| tény | T | B5 |
| lehet | K | B6 |
| talán | K | B6 |
| véletlen | K | B6 |
| akaratlan | K | B6 |
| önkéntelen | K | B6 |
| előreláthatatlan | K | B6 |
| öntudatlan | K | B6 |
| meggondolatlan | K | B6 |
| nagyon | K | B6 |
| időnként | K | B6 |
| megközelítőleg | K | B6 |
| megközelítően | K | B6 |
| kevés híján | K | B6 |
| majdhogy nem | K | B6 |
| szinte | K | B6 |
| borzalmasan | K | B6 |
| csaknem | K | B6 |
| talán | K | B6 |
| alkalmanként | K | B6 |
| esetlegesen | K | B6 |
| történetesen | K | B6 |
| tán | K | B6 |
| esetleg | K | B6 |
| netán | K | B6 |
| éppenséggel | K | B6 |
| valószínűleg | K | B6 |
| netalántán | K | B6 |
| meglehet | K | B6 |
| vélelmezhetően | K | B6 |
| becslés | K | B6 |
| saccolás | K | B6 |
| rettentően | K | B6 |
| azt lehet mondani | K | B6 |
| ritkán | K | B6 |
| mintegy | K | B6 |
| jóformán | K | B6 |
| majdnem | K | B6 |
| körülbelül | K | B6 |
| hit | K | B6 |
| nagyjából | K | B6 |
| különös módon | K | B6 |
| különös | K | B6 |
| leírhatatlanul | K | B6 |
| feltűnően | K | B6 |
| úgyszólván | K | B6 |
| kimondhatatlanul | K | B6 |
| hihetetlenül | K | B6 |
| túlontúl | K | B6 |
| durván számítva | K | B6 |
| durván | K | B6 |
| kvázi | K | B6 |
| épphogy nem | K | B6 |
| hozzávetőleg | K | B6 |
| némileg | K | B6 |
| hozzávetőlegesen | K | B6 |
| úgy-ahogy | K | B6 |
| kritik | K | B7 |
| reproduk | K | B7 |
| jócskán | K | B7 |
| kevese | K | B7 |
| rendkívül | K | B7 |
| irgalmatlanul | K | B7 |
| szerfelett | K | B7 |
| Fölöttébb | K | B7 |
| meglehetősen | K | B7 |
| felettébb | K | B7 |
| erőteljesen | K | B7 |
| bb | K | B7 |
| leg | K | B7 |
| hangsor | K | B7 |

1. számú melléklet: makró

Sub csere()

'csere Makró

'

' Billentyűparancs: Ctrl+q

'

Sheets("Munka2").Select

Dim gyakorisag As Integer

Dim attributum As String

For i = 1 To 492

gyakorisag = 0

attributum = Sheets("Munka1").Cells(3, i + 1)

gyakorisag = Application.CountIf(ActiveSheet.Range("A1:H22222"), "\*" & attributum & "\*")

Sheets("Munka1").Cells(4, i + 1) = gyakorisag

Next i

Sheets("Munka1").Select

End Sub

1. számú melléklet:

Mind a 18 TDK-dolgozatot csoportonkénti összesített találati eredményét egy táblázatban rögzítettük az alap TDK statisztikai mutatókkal együtt (oldalak száma, szavak száma, karakterek száma szóközzel, karakterek száma szóköz nélkül, stb). Ebből a táblázatból készült egy hasonlóságelemzés. Ahol az 1.000.000-s értéket választottuk meg a viszonyítási számnak, és ehhez az értékhez való eltérést néztük meg az egyes TDK-dolgozatok esetén. A kiválasztott dolgozatok között van olyan, ami myx-team lektorálta, s egy részét pedig olyan személy kinek nincs köze a myx-teamhez. Humán és real témájú dolgozatok egyaránt szerepelnek az elemzésben.

A minősítés szempontjából leginkább meghatározó csoportok:

* igazolt, bizonyított feltételezés
* tudományos kifejezések
* gyakoriságra utaló szavak (időbeli)

Az elkészült táblázatban minden sor adatait elosztottuk a sorban szereplő TDK-dolgozat karakterei számával szóközökkel együtt. A kapott eredményt 1000-rel szoroztuk, így megkaptuk, hogy az adott csoport a TDK-dolgozatnak hány ezrednyi részét teszi ki. Grafikonon szemléltettük milyen szórás mutatkozott meg a vizsgált dolgozatok becsült minősítése és a készítésen dátuma között. A vizsgált dolgozatok fele nem éri el a kívánt TDK minőségi szintet tartalmilag.