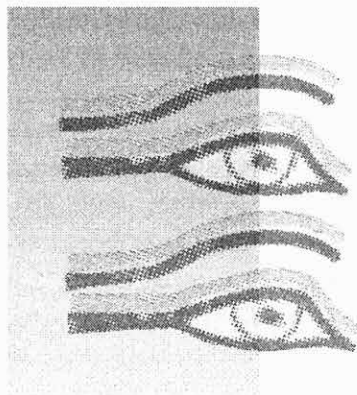


A kivonatkészítés sajátosságai egy felmérés adatainak tükrében*



LENGYELNÉ MOLNÁR Tünde

Kutatásaim egy automatikus kivonatoló program tervezésére irányultak, melynek nélkülözhetetlen előfeltételét jelentette az emberi kivonatkészítés sajátosságainak vizsgálata. Ennek megismerésére készítettem egy felmérést, melyben két különböző témájú szakcikk kivonatolására kértem fel több mint 250 főiskolai és egyetemi hallgatót, valamint könyvtáros szakembereket, továbbá kontrollként magyar szakos, szövegekkel bánti, azokat értelmezni tudó hallgatókat.

A felmérés alapjául szolgáló egy-egy cikket a *Könyvtári Figyelő*ből¹, illetve a *Tudományos és Műszaki Tájékoztatósból*² választottam.

A felmérésben résztvevőktől azt kértem, hogy készítsék el mindkét cikk kivonatát, annak leglényegesebb mondatainak megjelölésével. Mivel a cikkeknek 20%-os tömörítését vártam el a felmérésben résztvevőktől, ezért mindkét cikk esetén 17 mondatot kellett megjelölniük. Még azt is kértem tőlük, hogy a megjelölt mondatokat tegyék fontossági sorrendbe, és adják meg, hogy a cikk hányadik legfontosabb mondatának

tartják. Így egy 1 és 17 közé eső számmal jelölték a rangsort. A felmérésben szereplő szakemberektől a kivonat mondatainak rangsorolását számítógépen, egy online kérdőív kitöltésével egybekötve kértem.

A hallgatókkal papíron végeztem el a kivonatkészítést, megkönnyítve számukra az áttekinthetőséget, de a megoldások számítógépes feldolgoása utólagos adatbevitellel történt.

A felmérésben résztvevők

A felmérésbe igyekeztem minél több olyan felsőoktatási intézményt bevonni, ahol informatikus könyvtáros képzés folyik. Az egyetemi hallgatók bevonása során sajnos nem volt lehetőségem teljes mintavételt alkalmazni, így az egyetemek közül: a Debreceni Egyetem, az Eötvös Loránd Tudományegyetem, valamint a Szegedi Tudományegyetem könyvtár-informatika szakos hallgatóit kértem fel közreműködésre. Ezen intézményeknél az adott tanszéken tanító kollégák

* A tanulmány a szerző *Automatikus referátumkészítés* című PhD-disszertációjának lapunk számára átdolgozott, rövidített fejezete.

gondoskodtak a kitöltetésről.³

A főiskolai hallgatók az Eszterházy Károly Főiskola informatikus-könyvtár szakos hallgatói, akik közül a nappali és távoktatásos hallgatók teljes⁴ létszámmal részt vettek a felmérésben.

Mivel szerettem volna adatokat gyűjteni a szakemberek által készített referátumok jellemzőiről is, ezért a diákokon kívül felkértem a KATALIST könyvtáros levező listát olvasó szakembereket, hogy látogassanak el egy adott weboldalra⁵, ahol néhány kérdésre kellett válaszolniuk, és a két cikk kivonatát elkészíteniük. Mivel erre csak kevesen vállalkoztak, ezért több könyvtáros szakembert magánlevélben kértem fel a kitöltésre.

A felmérésben arra is kíváncsi voltam, mennyire fontos a szaktudás a kivonatok készítésekor, ezért egy kontroll csoportot is bevontam a felmérésbe: az Eszterházy Károly Főiskola magyar szakos hallgatóit. Azért kértem fel őket is a kivonatok elkészítésére, mert úgy gondoltam, ők biztosan rendelkeznek jártassággal különböző hosszúságú és témájú szövegek tömörítésében, a művek lényegének kiemelésében, de nem rendelkeznek könyvtártudományi szakismeretekkel, így alapot adnak a szaktudás és a kivonatkészítésben való jártasság értékének elemzésére. A végső minta 336 főből állt, megoszlásuk a következő volt:

ELTE	20 fő
Szegedi Tudományegyetem	31 fő
Debreceni Egyetem	5 fő
Eszterházy Károly Főiskola (magyar szak)	48 fő
Eszterházy Károly Főiskola (informatikus-könyvtáros szak)	180 fő
Szakemberek	52 fő
<i>Összesen:</i>	<i>336 fő</i>

A felmérés módja

A felmérés anyagát minden esetben tanóra alatt töltötték ki a hallgatók, kivéve a Debreceni Egyetem hallgatóit, akik otthon végezték a ki-

töltést. A hallgatók nem kaptak időlimitet a kitöltésre, de a tanóra kerete egyfajta korlátozásként jelent meg.

A szakemberektől, valamint az Eszterházy Károly Főiskola informatikus könyvtáros szakos hallgatóitól a kitöltő személyére vonatkozó néhány kérdésen túl a lényeges mondatok kiválasztását és rangsorolását kértem, az egyetemek hallgatóitól, valamint a magyar szakos hallgatóktól a fenti feladatokon kívül kértem, hogy jelöljék meg az általuk hasznosnak, relevánsnak ítélt szavakat is.

Eredmények

A következőkben a kapott értékeket előbb cikkenként vizsgáljuk, majd összevetjük egymással az eredményeket.

A „KOLTAY Tibor: Szöveg, információ, relevancia: néhány adalék a témakörhöz” c. cikkből készült kivonatok elemzése

Informatikus-könyvtáros hallgatók válaszai

A felmérésben résztvevő informatikus-könyvtáros szakos hallgatók 67,8%-a főiskolai szintű képzésben vesz részt, a 32,2%-uk jár egyetemi szintű képzésre.

A képzésben résztvevők átlag életkorának elemzéséből az derül ki, hogy a BA-szintű nappali képzésben inkább a középiskolát nemrég elhagyók vesznek részt (20 év alatti átlagéletkorukkal), míg az ELTE hallgatóinak közel 28 éves átlagéletkora azt tükrözi, hogy a felmérésben résztvevők 45%-a esti vagy levelező képzésben tanul. A Szegedi és Debreceni Tudományegyetemről résztvevő egyetemi hallgatók mindegyike nappali tagozatos, ami 23,0 és 22,8 éves átlagéletkorukban is megjelenik. A főiskolai szintű képzésben résztvevő szegedi hallgatók nappali tagozatosak (21,84 éves átlaggal), az Eszterházy Károly Főiskola felmérésében résztvevő hallgatóknak 60%-a távoktatási képzésben vesz részt, akiknek 26,57 éves átlagéletkora 7 évvel magasabb, mint a nappali tagozatos hallgatóké.

A megjelölt mondatok elemzése

Az első cikk kivonatát 208 informatikus-könyvtáros hallgató készítette el, ami 28 fővel kevesebb a minta létszámánál. A mondatok több mint 98%-a került megjelölésre valamely kitöltő személynél, azaz a felmérésben résztvevők által megjelölt mondatok széles skálán mozognak. Ez a spektrum jelentősen csökkenthető, ha kivesszük azokat a mondatokat, melyeket a kitöltők kevesebb, mint 10%-a jelölt meg. Így az eredeti mondatszámot 43%-kal csökkenthetjük. Ez azt jelenti, hogy azon mondatok, melyeket a kitöltők több mint 10%-a relevánsnak ítélt a szöveg feléből kerül ki. A további elemzések már nem ilyen látványosak.

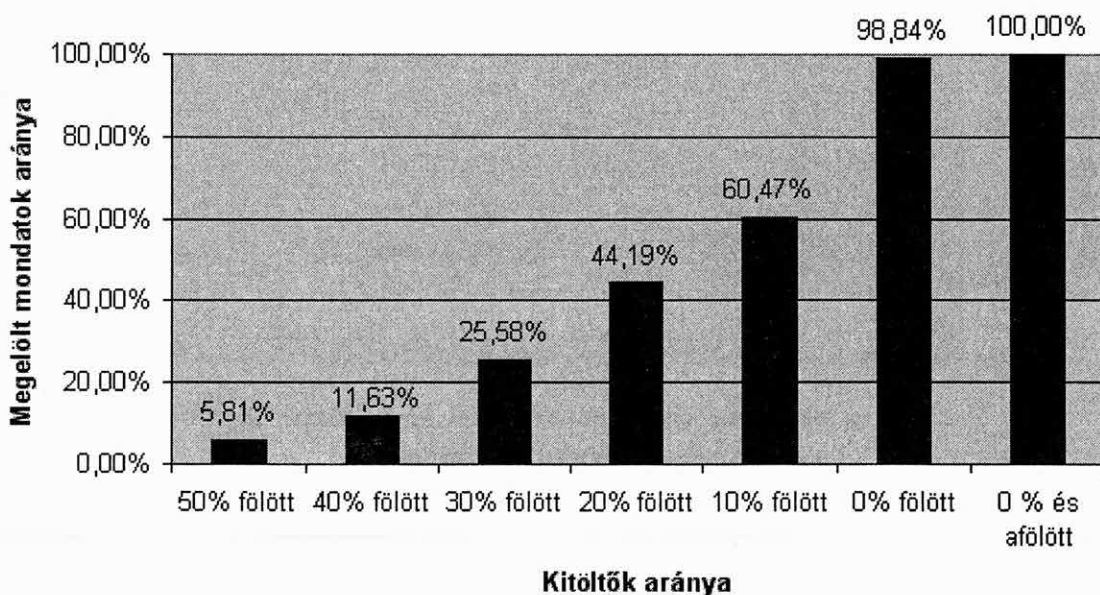
A legtöbb jelölést kapott mondatok vizsgálatakor kiderül, hogy 5 olyan mondata van a cikknek – az összes 86 mondat közül –, melyet a kitöltők több mint fele megjelölt valamilyen helyezéssel. Az 1. ábráról leolvasható, hogy ez a szöveg 5,81 %-át teszi ki.

A diagramról leolvashatjuk azt is, hogy a kitöltők több mint 30%-a megjelölte a szövegnek a 25,58%-át. Ebből már le is vonhatjuk azt a következtetést, hogy a szöveg 20%-os kivonatának mondatai adottak? A kitöltők 30%-a által megjelölt mondatok ugyanis kiteszik ezt a szöveganyagot. Ez a következtetés azért nem vonható le, mert nem hagyhatjuk figyelmen kívül a mondatok helyezéseit. Melyik mondat fontosabb: amelyet sokan megjelöltek, de rangsorukban csak az utolsó helyre tettek, vagy az a mondat, melyet fele annyian jelöltek meg, de előkelő helyezést adtak neki? A probléma kiküszöbölésére egy súlyozással történő vizsgálat bevezetésére volt szükség. Az eredmények megtekintése előtt érdemes még egy kis figyelmet szentelni a rangsoroknak, és összevetni a fentebb ismertetett eredményekkel.

A mondatok rangsorolásának elemzését kezdjük a legrelevánsabbal, azaz első helyre rangsorolt mondatok vizsgálatával (2. ábra). Két mondatot érdemes kiemelni:

1. ábra

A megjelölt mondatok alakulása informatikus-könyvtáros szakos hallgatók esetén



- a kitöltők 40%-a a cikk legelső mondatát tartotta a legrelevánsabbnak,
- ezt követte a cikk harmadik mondata, melyet a kitöltők 16,35%-a tartotta a leglényegesebbnek.

Látható, hogy a fenti mondatok a kitöltők fele által megjelölt első helyet foglalják magukban. A többiek első helyre tett mondatai 30 mondat között oszlanak el.

Ha bővítjük a kört és megnézzük mely mondatok kapták az első három hely valamelyikét, a kép nem változik sokat. A szöveg

- első mondatát a kitöltők 45,6%-a
- a harmadik mondatát a kitöltők 34,13%-a
- 2. és 14. mondatát a kitöltők 17,79%-a tette az első három helyezés valamelyikére.

A cikk első három mondatának mindegyike 1–1 önálló bekezdést is jelentett. Az első mondatok

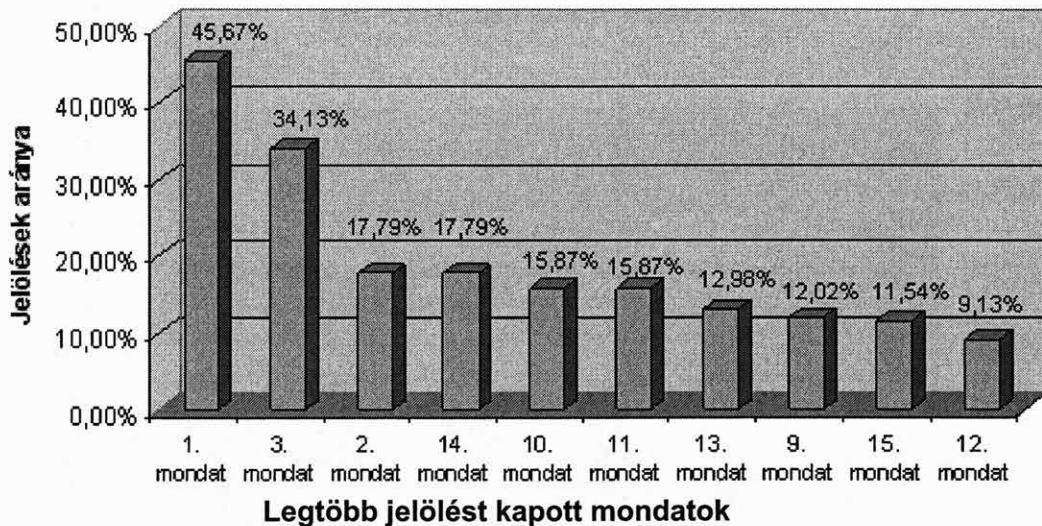
bekerülése a releváns mondatok közé nem meglepő. A tartalomelemző eljárások közül több is részletesen kitér az első bekezdés fontosságára, mivel a szerző itt vezeti be mondanivalóját, amit rendszerint olyan formában tesz meg, hogy ad egy összefoglaló gondolatsort a cikk tartalmáról, felsorolva a tárgyalt leglényegesebb témaköröket.

A teljesség kedvéért még néhány mondatot be kell vonni az értékelésbe, hiszen a 10. és 11. mondatot is megjelölte a kitöltők több mint 15%-a, továbbá még három mondat van, melyet a kitöltők több, mint 10%-a az első három hely valamelyikére rangsorolt.

Tekintsük meg annak a tíz mondatnak a sorszámtát, melyeket a legtöbben tettek az első három legfontosabb mondat közé:

2.ábra

Első három helyre rangsorolt mondatok elemzése informatikus-könyvtáros hallgatók esetén



Az első bekezdéseket súlyozó elméletek ugyanilyen fontosnak tartják az utolsó bekezdést is, mivel ott a szerző összegzi a cikkben foglaltakat, felsorolja, majd lezárja az elért eredményeket. A diagramon szereplő adatok alapján látható, hogy az utolsó bekezdés mondatai nem szerepelnek az első három helyre rangsorolt mondatok között, sőt az első három helyre a cikk első 15 mondatának valamelyike került a kitöltők több mint 60%-ánál.

Tegyük teljesebbé az elemzést egy súlyozás bevezetésével, melynél a helyezések megszámlálását súlyozott pontszám bevezetésével helyettesíttem. Az első helyre tett mondatot n -es súllyal vettem figyelembe, ahol „ n ” a megjelölhető mondatok számát jelenti, azaz a legnagyobb rangsorbeli elemet, majd a második helyre rangsorolt mondatok $n-1$ -es súllyal látam el, az utolsó helyre rangsorolt mondatokat pedig 1-es súllyal vettem figyelembe. Ezzel az eljárással kiküszöbölhető az a probléma, hogy melyik számít relevánsabb mondatnak, amit 10 ember első helyre tesz, vagy melyet 25 ember a

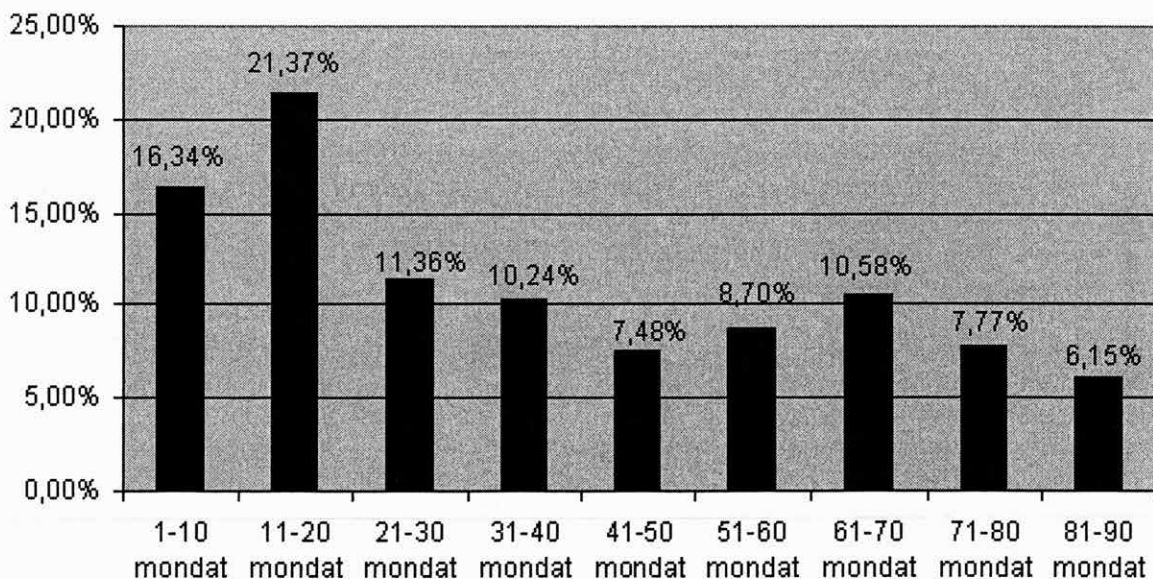
10. helyre rangsorol? Összesítettem a mintában szereplő személyek által adott jelölések alapján a mondatokhoz tartozó súlyozott értékeket, így a magasabb pontszámmal rendelkező mondatok megelőzik az alacsonyabb súlyozott értéket kapott mondatokat.

Ennek eredményeként a legrelevánsabbnak tartott mondat magasan a cikk első mondata, melyet a 3. mondat követ. Összegezve: a cikk első harmadából került ki a legtöbb mondat, a szöveg első harmadán túli részből összesen 3 mondatot fedezhetünk fel.

Az egyetemista könyvtár-informatikus hallgatóknál a felmérés részét képezte a releváns szavak megjelölése is. Sem mennyiségi korlát, sem rangsorolási feladata nem volt a hallgatóknak, csupán aláhúzással kellett jelölniük az általuk lényegesnek tartott szavakat.

Érdeemes megvizsgálni a szavak elemzése során készített kimutatást, melyből kiderül, hogy a megjelölt szavak a szöveg mely mondataiból származnak.⁶ Összhangban van-e ez a releváns mondatok megjelölésével?

3.ábra
Megjelölt szavak elhelyezkedése



A kivonat mondatainak 82,35%-a származik a szöveg első harmadából. A megjelölt hasznos szavaknak 44,5 %-a származik a szöveg ugyan-ezen részéből. Az összes mondathoz rendelt súlyozott értékek, valamint a mondatokban megjelölt kulcsszavak összevetése, tehát szoros kapcsolatra utal. Mindkét cikk esetén valamivel 0,8 feletti korrelációs értékek álltak elő, melyek az összes mondatot alapul véve, megerősítik a szoros kapcsolatot a két eredmény között.

A szakemberek válaszainak elemzése

A szakemberek számára készült online kérdőívet és kivonatot 52 fő töltötte ki. A kitöltők átlagos életkora 39,17 év, melyhez alacsony szórás érték párosult: 9,06 év. A legfiatalabb kitöltő 25 éves, a legidősebb 55 éves volt.

A megjelölt mondatok elemzése

A szakemberek válaszainak elemzésekor az első szembetűnő adat, hogy nagyon kevesen jelölték meg a kért 17 mondatot. Az első cikk kivonatát

48 fő készítette el, és átlagosan 10,35 mondatot jelöltek meg, míg ez az érték az informatikus-könyvtáros hallgatóknál 16,95 mondat.⁷

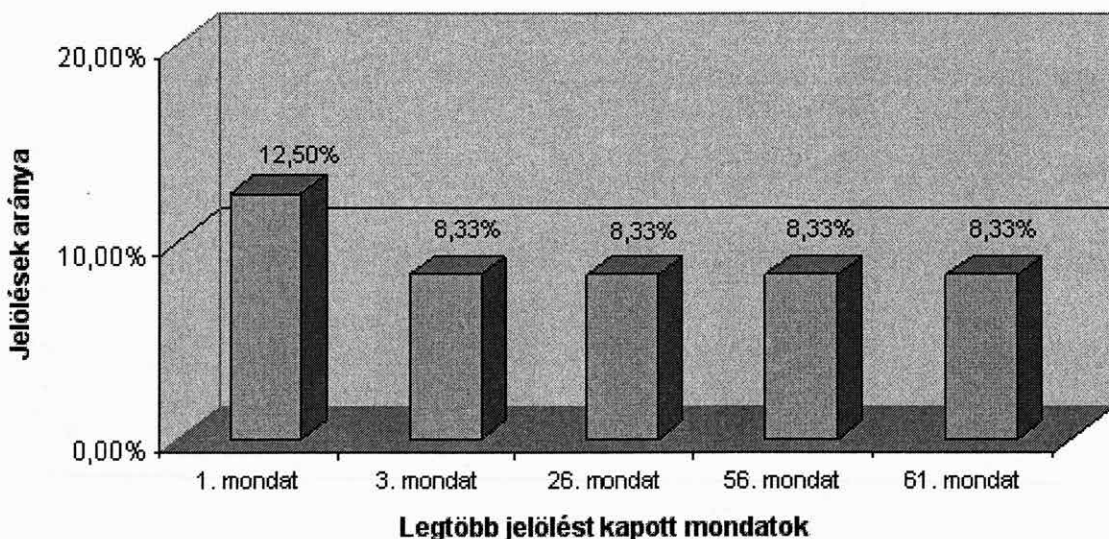
Ezzel szemben a mondatok eloszlása lefedi a teljes spektrumot, nincs olyan mondata a szövegnek, melyet senki sem jelölt meg.

A legtöbb jelölést kapott mondat is csak a kitöltők 25%-ának jelölését tudhatta magáénak, míg az egyetemista informatikus-könyvtáros hallgatók esetén az első mondatot 70,77%-uk jelölte meg. A főiskolai informatikus-könyvtáros hallgatóknál a 31. mondat kapta a legtöbb jelölést, melyet a kitöltők 66,43%-a tartott relevánsnak. A szakembereknél a 14. mondat kapta a legtöbb jelölést, mely az egyetemistáknál a negyedik, a főiskolásoknál pedig a nyolcadik helyen állt a jelölések számát alapul véve.

A legtöbb első helyezést a szakembereknél is az első mondat kapta, ezután azonban a hallgatói listától eltérő kép fogad minket. Az első három helyezést alapul véve a szakembereknél a 6. mondatról kezdve 15 mondat azonos számú jelölést kapott. Nézzük meg diagramon (4. ábra) ábrázolva az első 5 mondatot!

4. ábra

A szakemberek által első három helyre rangsorolt mondatok elemzése



Az első mondaton túl csupán a 3. mondat található meg az egyetemista és főiskolás hallgatóknál is, az egyetemista hallgatóknál még a 26. mondat szerepel a listában. Kibővítve a kört a 15 azonos jelölést kapott mondattal, a szakemberek által első három helyre tett mondatok közül még két mondat található meg a hallgatók hasonló kimutatásában.

Összegezve elmondható: míg a hallgatóknál az első 15 mondat a domináns, addig a szakembereknél a mondatok szövegen belüli elhelyezkedése sokkal egyenletesebb.

Nézzük meg, mely mondatok képeznék a 20%-os kivonatot a szakemberek válaszait alapul véve, súlyozva a jelöléseiket, majd a kapott eredményt összehasonlítjuk az egyetemista és a főiskolai hallgatók kivonatával.

Mondatok sorrendje	Szakemberek	Informatikus könyvtáros Egyetemista hallgatók	Informatikus könyvtáros Főiskolai hallgatók
1	1. mondat	1. mondat	1. mondat
2	2. mondat	14. mondat	3. mondat
3	31. mondat	26. mondat	31. mondat
4	3. mondat	15. mondat	14. mondat
5	14. mondat	11. mondat	26. mondat
6	16. mondat	3. mondat	24. mondat
7	67. mondat	31. mondat	37. mondat
8	12. mondat	50. mondat	15. mondat
9	65. mondat	24. mondat	50. mondat
10	55. mondat	18. mondat	23. mondat
11	32. mondat	2. mondat	13. mondat
12	26. mondat	10. mondat	10. mondat
13	51. mondat	13. mondat	16. mondat
14	70. mondat	9. mondat	11. mondat
15	76. mondat	16. mondat	19. mondat
16	42. mondat	12. mondat	18. mondat
17	47. mondat	37. mondat	9. mondat

A szakemberek 17 mondatos kivonatából az egyetemista hallgatók kivonatában 8 mondat szerepel, míg a főiskolai hallgatók kivonatával 6 mondatban egyezik meg. Ez 47%-os, illetve 35%-os átfedést jelent.

A kapott értékek arra utalnak, hogy a szakemberek releváns mondatkiválasztása jelentősen eltér az informatikus-könyvtáros hallgatók mondatkiválasztásától.

Az egyetemista és főiskolás hallgatók esetén az összes 86 mondathoz rendelt súlyozott pont-

számok összevetésekor kapott 0,9003-es korreláció nagyon szoros pozitív értéket mutat, amit úgy értelmezhetünk, ha egy mondat az egyik csoportnál magas értéket kapott, akkor a másik csoport tagjai is hasonlóan magas helyezéssel látták el az adott mondatot. Ezzel szemben a szakemberekkel történő összevetés során nem fedezhető fel szoros kapcsolat, bármely csoport is áll a másik oldalon. Az egyetemisták súlyozott pontszámai és a szakemberek válaszai 0,4862-es korrelációs értéket képeznek, mely laza, de

pozitív összefüggést mutat. A főiskolásokkal összehasonlítva a szakemberek válaszait, még lazább, pozitív összefüggést kapunk 0,4452-es korrelációs értékkel.

Ez az eredmény arra utal, hogy a szakemberek kivonatkészítése jelentősen eltér a hallgatók releváns mondatkiválasztásától. Az eddigi adatok tükrében az informatikus-könyvtáros hallgatók a szöveg elejét részesítik előnyben, míg a szakemberek a teljes korpuszból merítik a releváns mondatokat.

A kontroll csoport válaszainak elemzése

Már a vizsgálat elején foglalkoztatott a kérdés, hogy van-e összefüggés a szakmai cikkek és a kivonatkészítő szaktudása között, vagy ez a tu-

dás helyettesíthető a kivonatkészítésben való jártassággal. Ezért vontam be a kísérletbe az Eszterházy Károly Főiskola magyar szakos napali tagozatos hallgatói közül 48 diákot.

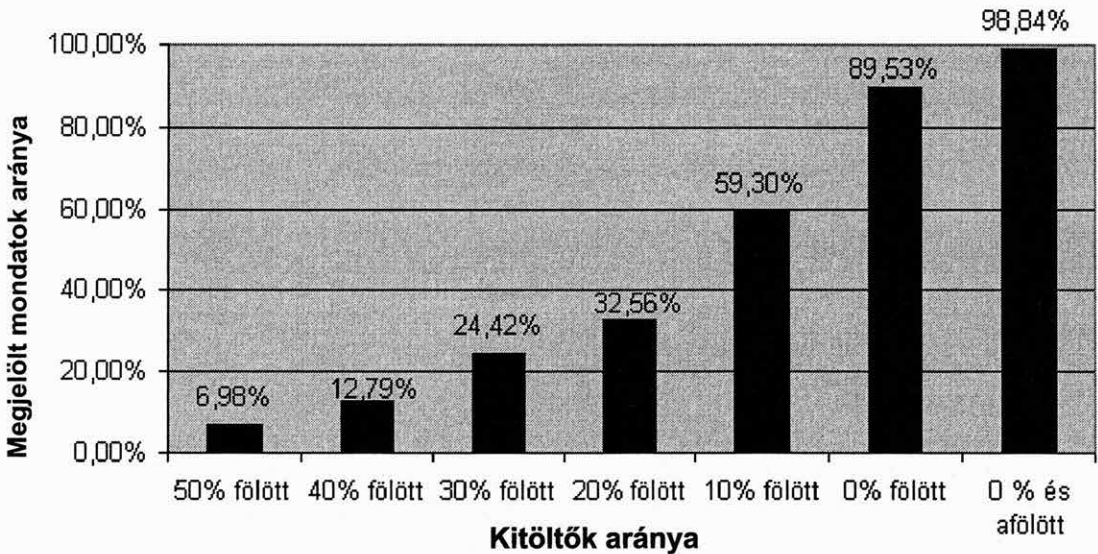
Nézzük meg, ők milyen rangsorokkal látták el Koltay Tibor cikkének mondatait.

A magyar szakos hallgatóknál 5 fő volt, aki csak a másik cikk kivonatát készítette el. A 43 kitöltő esetén azonban a másik három csoporthoz viszonyítva elég magas a meg nem jelölt mondatok száma. A 86 mondatból 8-at senki sem jelölt meg. A szakembereknél nincs olyan mondat, melyet senki sem rangsorolt volna, míg az egyetemistáknál két ilyen mondat van, a főiskolásoknál pedig három.

Így a magyar szakos hallgatók által megjelölt mondatok megoszlása a következő módon alakult.

5.ábra

A megjelölt mondatok alakulása kontroll csoport esetén



A legelső mondatot a kitöltők 74,42%-a jelölte meg, melyet a 11. mondat követ 62,79%-os jelöléssel. A magyar szakos hallgatók esetén kicsit magasabb a 40% fölötti jelölést kapott mondatok aránya, mint a többi csoportnál, mely érték a tizenegy mondatával az összes mondat 12,79%-át teszi ki.

Ez a két tényező, mely szerint kevesebb mondat között oszlanak meg a jelölések, illetve, hogy több mondat kapott magas számú jelölést, arra utal, hogy a magyar szakos hallgatók válaszaiban nagyobb az egyezés, nem szóródnak olyan széles spektrumon a válaszok, mint a másik három csoportnál.

Nézzünk meg egy összefoglaló táblázatot a megjelölt mondatok alakulásáról a négy minta esetén.

Mondatok által kapott jelölések aránya	Mondatok száma			
	<i>Szakemberek</i>	<i>Egyetemista hallgatók</i>	<i>Főiskolás hallgatók</i>	<i>Magyar szakos hallgatók</i>
0%	0,00%	2,33%	3,49%	9,30%
0,1% és 10 % között	24,42%	32,56%	38,37%	30,23%
10,1% és 20 % között	66,28%	25,58%	13,95%	26,74%
20,1% és 30 % között	9,30%	11,63%	17,44%	8,14%
30,1% és 40 % között	0,00%	16,28%	16,28%	11,63%
40,1% és 50 % között	0,00%	6,98%	4,65%	5,81%
50,1% és 60 % között	0,00%	3,49%	3,49%	5,81%
60,1% és 70 % között	0,00%	0,00%	2,33%	1,16%
70,1% és 80 % között	0,00%	1,16%	0,00%	1,16%

A szakemberek jelölési technikája tér el legjelentősebben a többi csoport rangsorolásaitól. Bár a magyar szakosoknál több a magas jelölést kapott mondat, a 20 és 40% közti középmezőny jóval alacsonyabb, mint az informatikus-könyvtáros hallgatóknál, az eltérés mégsem mondható jelentősnek. Nézzük meg a következő korrelációs mátrixot, melynek alapja, hogy a kitöltők hány százaléka jelölte meg az adott mondatot (mind a 86 mondat adatát összevetve a 4 mintacsoportban).

Korrelációs mátrix a megjelölt mondatok alapján	Szakemberek	Egyetemista hallgatók	Főiskolás hallgatók	Magyar szakos hallgatók
Szakemberek	1			
Egyetemista hallgatók	0,4442	1		
Főiskolás hallgatók	0,3965	0,8876	1	
Magyar szakos hallgatók	0,2852	0,8260	0,8089	1

A korrelációs értékek azt mutatják, hogy a szakemberek által megjelölt mondatok alakulása egyetlen másik csoporttal sincs korrelációs viszonyban, illetőleg az egyetemistákkal van egy gyenge, kimutathatóan pozitív korrelációs viszony, ezzel szemben a hallgatók jelölései szaktól függetlenül szoros pozitív korrelációs viszonyban állnak 0,8 fölötti értékükkel. A legerősebb összhang az informatikus-könyvtáros egyetemista és főiskolás hallgatók között van.

Nézzük meg hogyan alakultak a jelölések a magyar szakos hallgatók csoportjában!

A legtöbb első helyezést az első mondat kapta, melyet a kitöltők 60,47%-a rangsorolt az első helyre. Ezt a 3. mondat követi, melyet már csak a kitöltők 11,63%-a jelölt meg legrelevánsabb mondatként. A magyar szakosok egységesebb jelölését támasztja alá az is, hogy náluk összesen csupán 10 mondat szerepelt az első helyezették között, míg a szakembereknél 32 olyan mondat

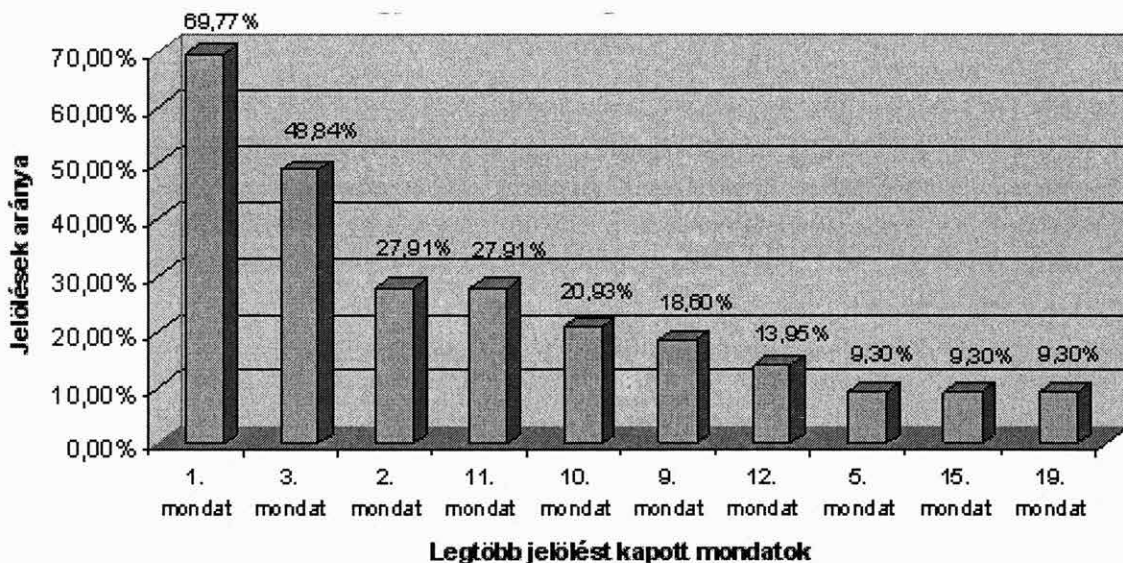
található, melyet valaki első helyre rangsorolt. Ez az érték az informatikus-könyvtáros hallgató egyetemista rétegénél 18, a főiskolások mintacsoportja esetén pedig 28 mondat.

Az első három helyre rangsorolt mondatok esetén is jóval szűkebb a megjelölt mondatok listája:

	Első három helyre rangsorolt mondatok száma
Magyar szakos hallgatók	23
Egyetemista hallgatók	44
Főiskolás hallgatók	57
Szakemberek	67

A megjelölt mondatok alakulása azonban már nem tér el olyan jelentős mértékben a szakhallgatók jelöléseitől:

6. ábra
Az első három helyre rangsorolt mondatok elemzése magyar szakos hallgatók esetén



Az első három helyre rangsorolt mondatoknál az első 15 mondat kapta a vezető szerepet, mint az eddigi hallgatói csoportok esetén is.

Mind a négy minta esetén az első helyezett a legelső mondat, melyet szintén mind a négy csoportnál a 3. mondat követ. Ezen túlmenően még említésre méltó, hogy a fenti diagram első 6 oszlopa teljesen megegyezik az informatikus-könyvtáros szakos egyetemisták által rangsorolt mondatokkal.⁸ A magyar szakosok tíz mondata közül az egyetemistáknál 8, a főiskolás hallga-

tóknál 7, a szakembereknél pedig 4 mondat található meg ugyanezen kimutatást alapul véve.

A leglényegesebbek természetesen a kivonatba kerülő mondatok.

Vizsgáljuk meg ezen a területen milyen eredményt kapunk, ha a magyar szakos hallgatók válaszait vesszük alapul!

A súlyozott pontszámok alapján a táblázatban található mondatok képezik a kivonat 17 mondatát, melyet hasonlítsunk most össze a másik három minta kivonatával!

Mondatok sorrendje	Magyar szakos hallgatók	Informatikus könyvtáros hallgatók		Szakemberek
		Egyetemista szakon	Főiskolai szakon	
1	1. mondat	1. mondat	1. mondat	1. mondat
2	11. mondat	14. mondat	3. mondat	2. mondat
3	3. mondat	26. mondat	31. mondat	31. mondat
4	15. mondat	15. mondat	14. mondat	3. mondat
5	12. mondat	11. mondat	26. mondat	14. mondat
6	19. mondat	3. mondat	24. mondat	16. mondat
7	10. mondat	31. mondat	37. mondat	67. mondat
8	23. mondat	50. mondat	15. mondat	12. mondat
9	2. mondat	24. mondat	50. mondat	65. mondat
10	31. mondat	18. mondat	23. mondat	55. mondat
11	14. mondat	2. mondat	13. mondat	32. mondat
12	9. mondat	10. mondat	10. mondat	26. mondat
13	26. mondat	13. mondat	16. mondat	51. mondat
14	16. mondat	9. mondat	11. mondat	70. mondat
15	13. mondat	16. mondat	19. mondat	76. mondat
16	24. mondat	12. mondat	18. mondat	42. mondat
17	50. mondat	37. mondat	9. mondat	47. mondat

Az informatikus-könyvtáros főiskolás és egyetemista hallgatók kivonatától csak két mondatban tér el a magyar szakosok kivonata. Ez 88%-os egyezőséget jelent. Érdekeség, hogy az egyetemisták és főiskolások kivonata is 2 mondatban tér el egymástól, a magyar szakosoknál pedig mind a 2–2 eltérő mondat megtalálható.

A szakemberek kivonata ezzel szemben lényegesen különböző, az azonos mondatok csupán a kivonat 47%-át teszik ki.

Nézzük meg a súlyozott pontszámokhoz tartozó korrelációs értékeket!

A következő értékek meghatározása során a 86 mondatához tartozó súlyozott pontszámokat vettem alapul a négy minta esetén:

Súlyozott pontszámokból képzett korrelációs mátrix				
	<i>Informatikus-könyvtáros egyetemista hallgatók</i>	<i>Informatikus-könyvtáros főiskolás hallgatók</i>	<i>Szakemberek</i>	<i>Magyar szakos hallgatók</i>
Informatikus-könyvtáros egyetemista hallgatók	1			
Informatikus-könyvtáros főiskolás hallgatók	0,9167	1		
Szakemberek	0,4863	0,4452	1	
Magyar szakos hallgatók	0,8716	0,8559	0,4414	1

A súlyozott pontszámokhoz tartozó korrelációs értékek alátámasztják az eddigi elemzéseket. A hallgatói csoportok között minden irányban szoros korrelációs viszony áll fenn, mely alapján, ha valamely hallgatói rétegnél az adott mondathoz alacsony vagy magas pontszám tartozik, akkor a többi hallgatói rétegnél is hasonlóan alacsony, vagy magas pontszám állt elő. A legszorosabb viszony az informatikus-könyvtáros egyetemista és főiskolás hallgatók között van 0,9167-es korrelációs értékkel, majd az egyetemista könyvtárosok eredményéhez áll a legközelebb a magyar szakos hallgatók adataiból nyert súlyozott pontoszlop 0,8716-os korrelációs érték alapján. A főiskolás informatikus-könyvtáros hallgatók és a magyar szakos hallgatók 17 mondatból álló kivonatára igaz, hogy csak két mondatban tér el, mégis amikor az összes mondatot vesszük alapul, akkor a köztük lévő viszony alacsonyabb korrelációt mutat, mint az egyetemistákkal való kapcsolat, de a 0,8559-es érték, még mindig szoros pozitív viszonyt mutat.

A szakemberekkel való összehasonlítás során jóval alacsonyabb értékeket láthatunk, melynek 0,4 fölötti számadatai ugyan még pozitív korrelációról tanúskodnak, de ez a viszony elég

laza kapcsolatot takar, mivel épphogy átlépi a kimutatható kapcsolat határát.

Kivételt az egyetemista hallgatók jelentenek, mivel ezen csoport rangsoraiból képzett súlyozott értékkel szorosabb 0,4863-as korreláció alakult ki, mely erősebb kapcsolatot jelent, mint a másik két mintával számított érték.

A magyar szakos hallgatók is részt vettek a kivonat előállításán túl a szöveg kulcsszavainak megjelölésében. A szöveg első felének előnyben részesítése nemcsak a kivonatkészítésre jellemző, hiszen a kivonatnak a 88,23%-a a szöveg első harmadának mondataiból állt össze, de a lényegesnek tartott szavak 50,17%-a is a szöveg ezen részéből került ki.

A későbbi vizsgálatok azt az eredményt hozták, hogy a hallgatók által megjelölt preferált szavak a szöveg első részéből származnak, bár a szöveg teljes területéről történt a szókiválasztás (mindkét cikk esetében).

Az összes mondathoz rendelt kulcsszavak és a rangsorolásból származó súlyozott értékek korrelációs elemzése szoros pozitív kapcsolatot mutat. Ez az érték a magyar szakos hallgatók esetén mindkét cikknél 0,7697.⁹

A szöveg	Koltay Tibor cikke esetén	Prokné Palik Mária cikke esetén
	<i>A kulcsszavak aránya</i>	<i>A kulcsszavak aránya</i>
<i>I. negyede</i>	45,62%	37,37%
<i>II. negyede</i>	21,69%	22,49%
<i>III. negyede</i>	15,05%	21,11%
<i>IV. negyede</i>	17,64%	19,03%

Az első cikk elemzése után azt mondhatjuk el, hogy a hallgatói csoportoknál nem jelent számottevő különbséget a szakmai tudás adta előny a kivonatkészítésben való jártassághoz képest. Azonban a hallgatók releváns mondatkiválasztási technikája jelentősen eltér a szakmai tapasztalatokkal rendelkező szakemberek kivonatkészítési eredményeitől.

Nézzük meg, hogy a másik cikk elemzése alátámasztja-e, vagy elveti ezt a végkövetkeztetést!

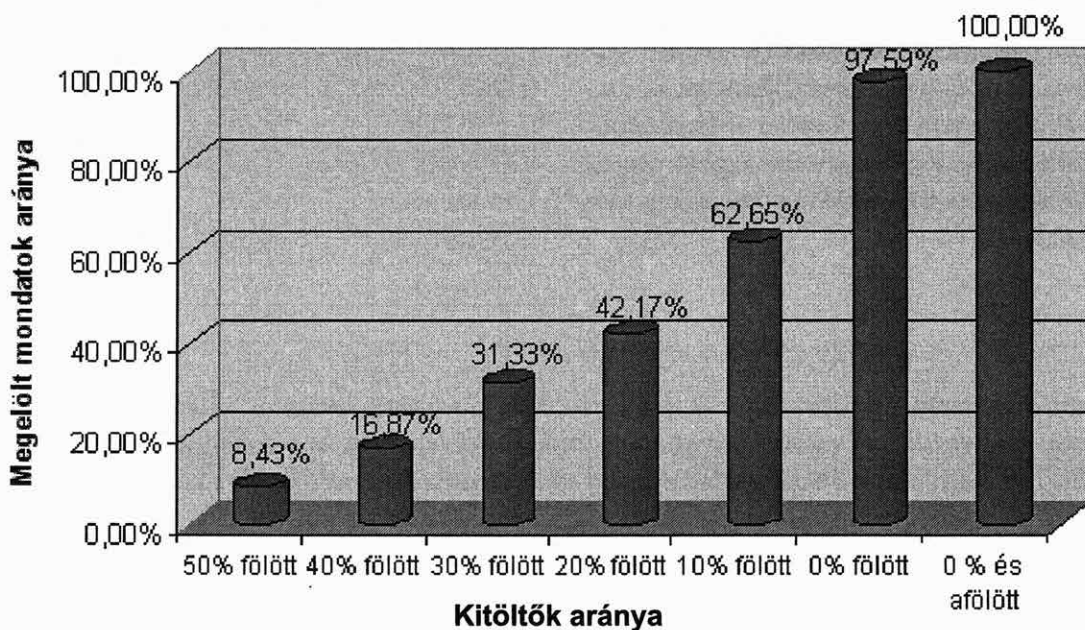
PROKNÉ Palik Mária: *A tartalmi feltárás problémái online könyvtári katalógusokban című cikkéből készült kivonatok elemzése*

Informatikus-könyvtáros hallgatók válaszai

A második cikk 83 mondatot tartalmaz, ebből csupán 2 olyan mondat van, melyet az informatikus-könyvtáros hallgatók közül senki sem jelölt meg. A mondatok 37,35%-a kevesebb, mint

a kitöltők 10%-ánál szerepel a rangsorolásban. A második cikk kitöltését a 236 informatikus-könyvtáros hallgatók közül 201 fő készítette el. Őket alapul véve, nézzük meg, hogy a megjelölt mondatok a kitöltők hány százalékának jelölésével rendelkeznek:

7. ábra
**A megjelölt mondatok alakulása
 informatikus-könyvtáros hallgatók esetén**



Két mondat van, melyet a kitöltők több, mint 70%-a megjelölt, ezen felül még 5 mondat jelölése kapott 50% feletti jelölést. A magas arányú jelölések száma több, mint az első cikknél tapasztalt érték, azonban a teljes gyakorisági táblázatot alapul véve hasonló jelölési technikát láthatunk a két cikk esetén.

Az első hely 30 mondat között oszlik meg. A legtöbben a legelső mondatot tartották a legrelevánsabbnak, melyet a kitöltők 32,34%-a rangsorolt a leglényegesebb mondatként. Ezt a második mondat követi a kitöltők 10,95%-ának első helyezését magának tudva, majd a sort a tizedik mondat folytatja a kitöltők 9,95%-ának első helyezéssel.

Ugyanezzel a három mondattal találkozunk akkor is, ha az első három helyre rangsorolt mondatokat számláljuk meg, és viszonyítjuk az ös-

zes informatikus-könyvtáros hallgató válaszához. Az első három helyre tett vezető tíz (azonos jelölések miatt tizenegy) mondat elemzésének eredményeként azt kapjuk, hogy ezen cikk esetén is túlmenően a szöveg elején lévő mondatok dominálnak, de emellett a szöveg legvégén lévő mondat is szerepel a tízes listán. Ez mindenképp eltér az első cikknél tapasztaltaktól, ahol a szöveg első 15 mondatán túl elhelyezkedő mondatok nem kaptak előkelő rangsorolásokat ugyan ezen mintacsoport esetén.

A megjelölések elemzésekor figyelembe kell venni az eredeti szöveg felépítését is. A cikk első bekezdése 6 mondatból áll. Mindenképp említésre méltó, hogy ebből a 6 mondatból csupán egy (a 3.) mondat nem szerepel az első három helyezésből képzett tízes listán. Ebből azt a következtetést vonhatjuk le, hogy a kitöltők

nagy százaléka az első bekezdés mondatait tette az első három hely valamelyikére.

Számszerűsítve az adatokat: a kitöltők 53,23%-a tette az első helyre az első 6 mondat valamelyikét, 41,79%-uk adta a második helyezést ezen 6 mondat egyikének, illetve a mintacsoport 26,37%-ánál a 6 mondat egyike birtokolja a harmadik helyezést.

A legtöbb jelölést kapott mondatok azonban egyáltalán nem az első bekezdésből kerültek ki, sokkal nagyobb hangsúly tevődik a szöveg második felére. A tíz legtöbb jelölést kapott mondat között három mondat a szöveg utolsó negyedéből származik.

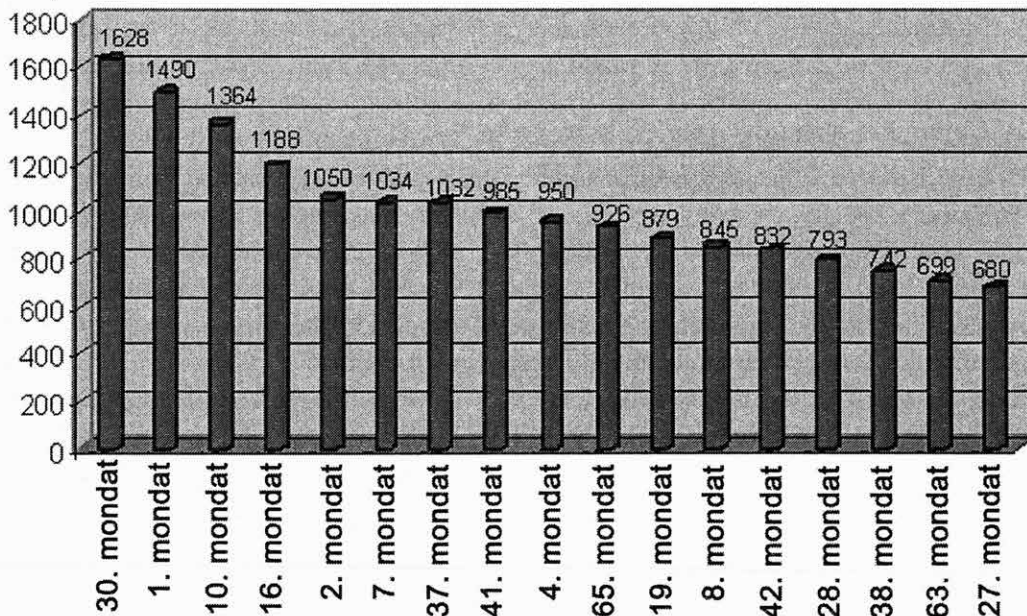
Legtöbb jelölést kapott mondatok		
Mondat sorszáma	A kitöltők hány százaléka jelölte meg?	Átlagos helyezés
30. mondat	77,11%	7,41
65. mondat	75,12%	12,46
81. mondat	57,71%	14,75
41. mondat	57,21%	9,39
63. mondat	55,72%	12,07
1. mondat	51,74%	3,67
37. mondat	51,24%	7,99
10. mondat	49,25%	4,30
42. mondat	49,25%	9,64
16. mondat	46,77%	5,32

Nézzük meg, mely mondatok képezik az informatikus-könyvtáros hallgatók válaszaiból előállt kivonatot! A súlyozott pontszámítás elve természetesen a második cikk esetén is megegyezik az eddig alkalmazott módszerrel. Ennek eredményeként a következő 17 mondat került a kivonatba:

8. ábra

A kitöltők által legrelevánsabbnak tartott mondatok informatikus-könyvtáros hallgatók esetén

súlyozott pontszám



A kivonatban a szöveg első bekezdésének hat mondata közül hármat találunk meg, illetve még három mondat (a 7., 8. és 10.) is a szöveg elejéről származik.

Ezen cikk esetén nem kap akkora szerepet a szöveg első negyede, mint a Koltay-cikk esetében. Hasonlítsuk össze a kivonatba került mondatok elhelyezkedését az eredeti szövegek alapján!

A szöveg	Koltay Tibor cikke esetén		Prokné Palik Mária cikke esetén	
	Mondatszám	A kivonat hány %-a?	Mondatszám	A kivonat hány %-a?
I. negyede	11	64,7%	8	47%
II. negyede	5	29,41%	6	35%
III. negyede	1	5,88	1	5,88%
IV. negyede	0	0	2	11,76%

Vessük össze ezt a kimutatást a megjelölt kulcsszavak elhelyezkedését mutató táblázattal:

A szöveg	Koltay Tibor cikke esetén	Prokné Palik Mária cikke esetén
	A kulcsszavak aránya	A kulcsszavak aránya
I. negyede	43,08%	34,75%
II. negyede	19,66%	22,41%
III. negyede	15,21%	24,40%
IV. negyede	22,05%	18,43%

A második cikknél a kulcsszavak elhelyezkedése sokkal egyenletesebb. Bár itt is a szöveg első negyedéből származik a legtöbb kiválasztott szó, azonban a szöveg többi részében található kifejezések sem lettek figyelmen kívül hagyva.

A szakemberek válaszainak elemzése

Az 52 kitöltő szakember közül mindenki elkészítette a kivonatát *Prokné Palik Mária* cikkének, igaz kevesen jelöltek meg 17 releváns mondatot (átlagosan 11,42 mondat került megjelölésre.) Az előző cikknél tapasztalt jelenség, mely szerint jóval kevesebb mondat van, amely magas számú jelölést ért el ennél a mintacsoportnál is felfedezhető. Egyetlen mondat van, amelyet nem jelölt meg senki: ez a főiskolás szakhallgatóknál a 2. mondat, az egyetemistáknál pedig az 5. mondat. Azonban a legtöbb jelölést kapott mondat is csak a kitöltők 28,85%-ánál került be a releváns mondatok közé, míg ez az érték az informatikus-könyvtáros hallgatók esetén 70, illetve 80% fölötti.

A legtöbb jelölést kapott mondat azonban meg-

egyeznek a hallgatók eredményével: a 30. mondat vezeti a listát a szakembereknél is. (A főiskolás hallgatóknál ez a második leggyakrabban megjelölt mondat, de csak 1%-kal kevesebb jelölést birtokolva, mint a vezető mondat, az egyetemisták esetében pedig ez a legtöbb jelölést szerzett mondat.) A szakembereknél még másik két mondatot kell megemlíteni, melyek a 30. mondattal azonos számú jelölést kaptak, ezek az 1. és a 7. mondatok.

A legtöbb első helyezést kapott mondatban is egyetértés van a három mintacsoport között, mindegyiknél a szöveg első mondata került az élre. (Bár a szakembereknek csak a 7,69%-a rangsorolta legrelevánsabb mondatként). Az ezt követő mondatok a szakembereknél már jelentős mértékben eltérnek a hallgatók véleményétől. Nem beszélhetünk egységes álláspontról, mivel az 52 szakember 37 különböző mondatot rangsorolt az első helyre. A főiskolások majd háromszoros kitöltő létszáma mellett 27 mondat került az első helyre, a hasonló számú egyetemisták legrelevánsabb mondatai 16 mondat között oszlanak meg.

Az első három helyezést alapul vevő listán 7 mondat van, mely kiemelhető, a következő 13 mondat azonos számú jelölést kapott.

A szakembereknél is kiemelkedő jelölést kaptak az első bekezdés mondatai. A bekezdés hat mondatából öt mondat kapta a legtöbb dobogós

helyezést. A hét mondatból négy található meg a főiskolás szakhallgatók hasonló elemzésében, és öt mondatban egyezik meg az egyetemista szakhallgatók jelölésével.

Nézzük meg a súlyozott pontszámok alapján előálló kivonat mondatait!

Mondatok sorrendje	Szakemberek	Informatikus-könyvtáros szakos egyetemista hallgatók	Informatikus-könyvtáros szakos főiskolás hallgatók
1	30. mondat	30. mondat	30. mondat
2	3. mondat	1. mondat	10. mondat
3	1. mondat	2. mondat	1. mondat
4	2. mondat	7. mondat	16. mondat
5	6. mondat	10. mondat	37. mondat
6	7. mondat	28. mondat	41. mondat
7	5. mondat	16. mondat	4. mondat
8	65. mondat	42. mondat	7. mondat
9	51. mondat	65. mondat	2. mondat
10	39. mondat	8. mondat	65. mondat
11	72. mondat	37. mondat	19. mondat
12	41. mondat	19. mondat	8. mondat
13	31. mondat	4. mondat	27. mondat
14	23. mondat	41. mondat	42. mondat
15	9. mondat	63. mondat	38. mondat
16	47. mondat	38. mondat	28. mondat
17	19. mondat	81. mondat	63. mondat

A szakemberek kivonata csak 7–7 mondatban egyezik meg a szakhallgatók kivonatával. Ez 41%-os egyezést jelent. A kapott érték összhangban van az előző cikknél kapott értékekkel, mert

ott 47%, és 35%-os volt az egyezés.

Érdeemes még megnézni, hogy a szakembereknél hogyan alakul a kivonat mondatainak elhelyezkedése az eredeti szöveghez viszonyítva.

A szöveg	Koltay Tibor cikke esetén		Prokné Palik Mária cikke esetén	
	Mondatszám	A kivonat hány %-a?	Mondatszám	A kivonat hány %-a?
I. negyede	6	35,29%	8	47,05%
II. negyede	4	23,53%	5	29,41%
III. negyede	3	17,64%	2	11,76%
IV. negyede	4	23,53%	2	11,76%

Az első cikk esetén egyenletesebb a releváns mondatkiválasztás, mint a szakhallgatók esetén, akik a szöveg elejét preferálták. A második cikk esetén viszont a releváns mondatok eloszlása megegyezik a szakhallgatóknál kapott adatokkal.

A kontroll csoport válaszainak elemzése

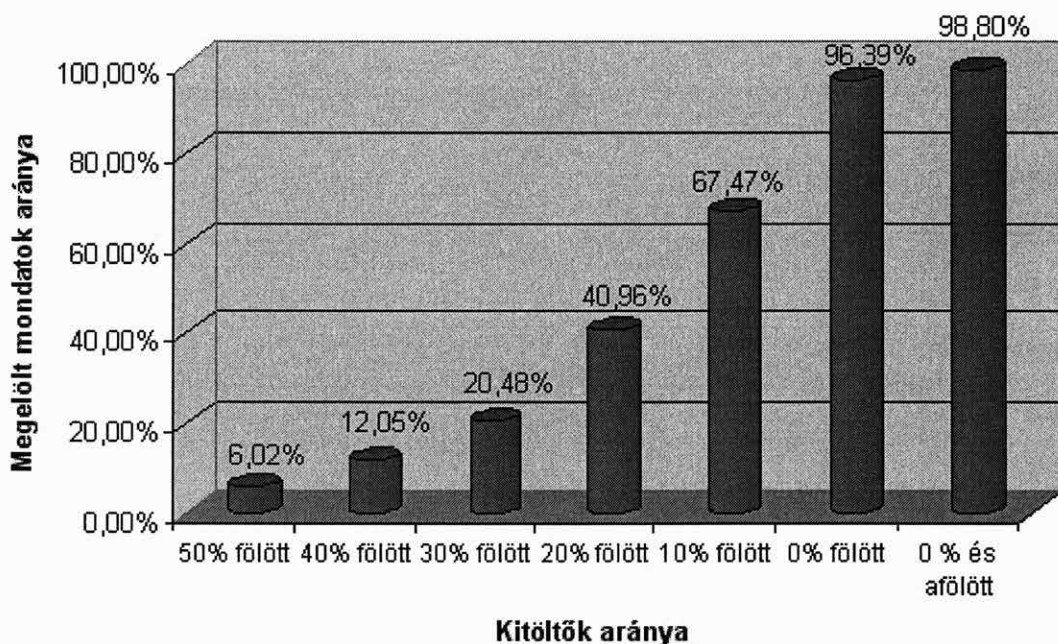
A magyar szakos hallgatók közül hat fő nem készítette el a második cikk kivonatát, így 42 fő rangsorolásának elemzésére van lehetőségünk. Két olyan mondata van a szövegnek melyet nem jelölt meg közülük senki, mely lényegesen alacsonyabb az előző cikknél tapasztalt 8 mondatos

értéktől, de nem tér el jelentősen a másik három mintacsoportban kapott eredménytől.¹⁰

A legtöbb jelölést az első mondat tudhatja magáénak, melyet a kitöltők több, mint 70%-a megjelölt. 10%-al kevesebben jelölték meg a második helyre szorult 30. mondatot. (Mely a szakembereknél és egyetemistáknál a legtöbb jelölést kapta, a főiskolásoknál szintén a második helyen van).

9. ábra

A megjelölt mondatok alakulása kontroll csoport esetén



A diagramról azt olvashatjuk le, hogy a magas jelölések elmaradnak a szakhallgatók jelöléseitől. A 30% feletti jelölések 12–15%-kal maradnak el a másik két hallgatói csoporttól, viszont a szakemberek magas arányú jelöléseit messze fölül múlják, hiszen náluk olyan mondatokkal nem is találkozunk, melyet a kitöltők több, mint 30%-a megjelölt volna. Koltay Tibor cikkénél nem volt ilyen jelentős eltérés a mintacsoportok között.

A legtöbb első helyezést az első mondat kapta, melyet a kitöltők közel 60%-a¹¹ tartott a legre-

levánsabbnak. Ebben mind a négy mintacsoport egyetért, és a folytatásban sem találunk eltérést. A magyar szakos hallgatóknál ezt a 2. majd a 10. mondat követi, mely sorrend teljesen megegyezik a főiskolás informatikus-könyvtáros hallgatók első helyre rangsorolt mondatainak sorrendjével, az egyetemista szakhallgatóknál is ez a két mondat folytatja a listát, csak fordított sorrendben, a szakembereknél azonban csak a tizedik helyre kerültek ezen mondatok¹² a legrelevánsabbnak tartott mondatok listáján.

A legnagyobb egyetértés a kitöltők között ebben

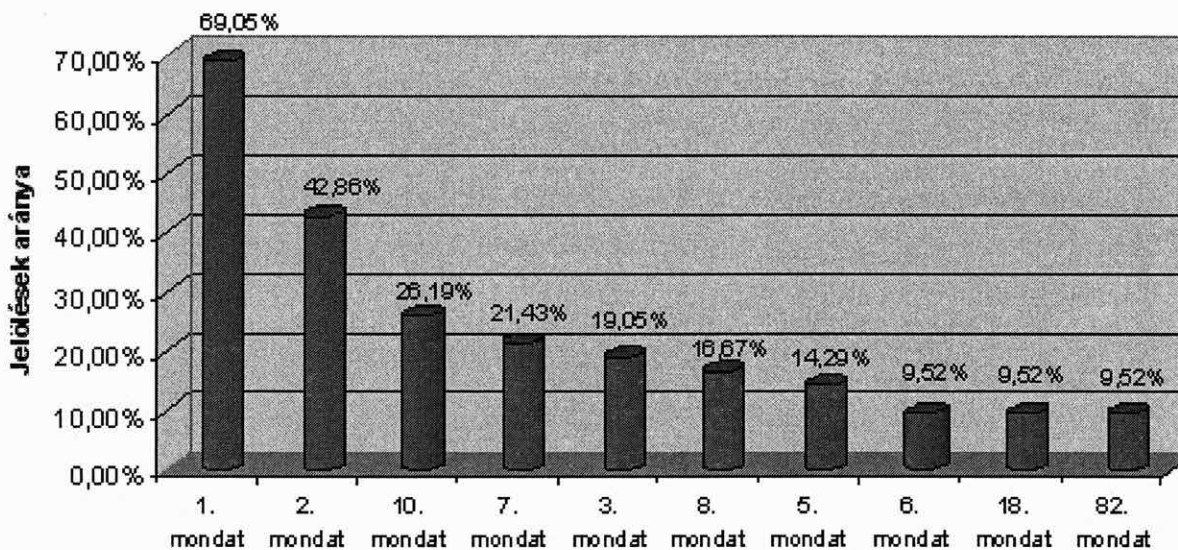
a csoportban tapasztalható, összesen tíz mondat került fel az első helyre rangsorolt mondatok közé.

Az első három rangsorbeli helyezés összesítése után is ugyanez a három mondat vezeti a listát, és az előző mintacsoportokhoz hasonlóan a legelső bekezdés mondatai közül csupán egy nem

került a listára. A magyar szakosoknál és a szakembereknél az első bekezdés mondatai közül a negyedik mondat maradt ki a preferált mondatok közül, míg az informatikus-könyvtáros hallgatóknál ez megtalálható, de a harmadik mondat kapott alacsony számú jelölést.

10. ábra

Az első három helyre rangsorolt mondatok elemzése a kontroll csoport esetén



Legtöbb jelölést kapott mondatok

A szöveg utolsó mondatainak kiemelése viszont a hallgatók körében jelent meg, a 82. (utolsó előtti) mondat a főiskolás szakhallgatóknál is, illetve a magyar szakos hallgatóknál is kiemelkedő számú dobogós helyezést kapott, míg az egyetemista hallgatók a 81. mondatot preferálták hasonló képen. A szakembereknél viszont a szöveg utolsó mondatai nem kerültek be egyetlen kimutatásba sem.

Most azt nézzük meg mely mondatok alkotják a kontroll csoport kivonatát (*ld. a következő oldalon lévő táblázatot!*)

A magyar szakos hallgatók kivonatában nem található a szöveg utolsó negyedéből származó mondat. A szakhallgatók kivonatától ugyanabban az öt mondatban tér el, illetve az egyetemisták kivonatától hat mondatos az eltérés. Ez a főiskolai szakos informatikus-könyvtáros

A kontroll csoport megállapításai

Mondatok sorrendje	Magyar szakos Hallgatók	Informatikus könyvtáros hallgatók		Szakemberek
		<i>Egyetemista szakon</i>	<i>Főiskolai szakon</i>	
1	1. mondat	30. mondat	30. mondat	30. mondat
2	2. mondat	1. mondat	10. mondat	3. mondat
3	7. mondat	2. mondat	1. mondat	1. mondat
4	10. mondat	7. mondat	16. mondat	2. mondat
5	30. mondat	10. mondat	37. mondat	6. mondat
6	8. mondat	28. mondat	41. mondat	7. mondat
7	16. mondat	16. mondat	4. mondat	5. mondat
8	19. mondat	42. mondat	7. mondat	65. mondat
9	42. mondat	65. mondat	2. mondat	51. mondat
10	3. mondat	8. mondat	65. mondat	39. mondat
11	37. mondat	37. mondat	19. mondat	72. mondat
12	5. mondat	19. mondat	8. mondat	41. mondat
13	23. mondat	4. mondat	27. mondat	31. mondat
14	18. mondat	41. mondat	42. mondat	23. mondat
15	27. mondat	63. mondat	38. mondat	9. mondat
16	41. mondat	38. mondat	28. mondat	47. mondat
17	36. mondat	81. mondat	63. mondat	19. mondat

hallgatók kivonatával több, mint 70%-os egyezőséget, az egyetemista hallgatók kivonatával pedig 64,7%-os átfedést jelent. A magyar szakos hallgatók kivonatának van a legtöbb közös mondata a szakemberek kivonatával, nyolc mondatban tér el attól, ami 52,94%-os egyezőséget jelent. (A szakemberek kivonata a szakhallgatókkal 47%-os, illetve 35%-os egyezést mutatott.)

Pontosabb képet kapunk a mintacsoportok mondatkiválasztási technikájának hasonlóságáról, ha a súlyozott pontszámokhoz tartozó korrelációs mátrixot elemezzük. A mátrix alapja a 83 mondathoz rendelt súlyozott pontszám a négy mintacsoport esetén.

Súlyozott pontszámokból képzett korrelációs mátrix				
	<i>Informatikus könyvtáros egyetemista hallgatók</i>	<i>Informatikus könyvtáros főiskolás hallgatók</i>	<i>Szakemberek</i>	<i>Magyar szakos hallgatók</i>
Informatikus könyvtáros egyetemista hallgatók	1			
Informatikus könyvtáros főiskolás hallgatók	0,9361	1		
Szakemberek	0,5261	0,4708	1	
Magyar szakos hallgatók	0,8781	0,8487	0,5758	1

Mind a négy csoport között pozitív korrelációs kapcsolat mutatható ki, melyek erősebbek, mint Koltay Tibor cikkénél. A legerősebb viszony most is az informatikus-könyvtáros hallgatói réteg között áll fenn, melynek 0,9361-es értéke nagyon hasonló gondolkodásmódot tükröz. Ezt követően ugyanúgy, mint az előző cikknél a magyar szakos hallgatók és az egyetemista szakhallgatók mondatkiválasztási technikája áll a legközelebb egymáshoz, melyet csak kicsivel alacsonyabb értékkel a főiskolás szakhallgatók és magyar szakosok kapcsolata követ. Mindkét esetben szoros pozitív korrelációról beszélhetünk 0,8 fölötti értékekkel.

A kivonatot elkészítő szakemberek mondat-megjelölési technikájához a magyar szakos hallgatók eredménye áll a legközelebb, mely korrelációnak 0,5758-as értéke egyáltalán nem elhanyagolható. Bár nem éri el a 0,7 fölötti szoros viszonyra utaló korrelációs értéket, de jóval magasabb, mint az előző cikknél kapott, a kimutatható kapcsolat határán lévő 0,4414-es korrelációs érték. Az informatikus-könyvtáros hallgatókkal is szorosabb a kapcsolat, mint a Koltay-cikk esetében: az egyetemista hallgatókkal 0,5261-es, a főiskolai szakhallgatókkal pedig 0,4708-as pozitív korrelációs kapcsolat áll fenn.

Összefoglalás

A két cikk elemzése során kapott értékek összhangban vannak, alátámasztják egymást.

Az informatikus-könyvtáros hallgatói réteg egyetemista és főiskolás szakra járó felosztását követő elemzéskor kapott releváns mondatok kiválasztási és rangsorolási technikája megdöbbentő hasonlóságot mutat. Ez alapján kijelenthetjük, hogy a kivonatolás nem függ a szakhallgatók képzésének fokától.

A magyar szakos hallgatók rangsorolási technikáját elemző súlyozott pontszámok is mindkét cikk esetén 0,8-as korrelációs szint fölötti egyezést mutatnak mindkét szakhallgatói csoporttal. Ennek következményeként elmondhatjuk, hogy szakmai cikkek kivonatolása során a szaktudás és a szaktudás nélküli, de a tömörítésben, lényegkiemelésben való jártasság nem eredményez számottevő eltérést a hallgatói csoportok esetén.

Jelentős eltérést tapasztalunk azonban a hallgatók és a szakemberek kivonatolása között. A hallgatók esetében a szöveg elején elhelyezkedő mondatok voltak a preferáltak, míg a szakemberek a kivonat mondatait a szöveg teljes spektrumából merítették.

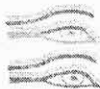
A hallgatói jelölések sokkal egységesebbek,

mint az eltérő szakmai háttérrel rendelkező, a szakma különböző szakterületeiről származó szakemberek rangsorolásai.

A felmérés célja a kivonatolás automatizálása során figyelembe veendő sajátosságok feltárása volt. Ennek eredményeként a legfontosabb megállapítás az, hogy a kivonatolás automatizálását célzó eljárásoknál mindenképp figyelembe kell venni a bevezető mondatok preferáltságát, mely mind a négy mintacsoportnál kimutatható volt. Másik eredmény, a mondatonkénti kulcsszavak és a rangsorolás során kapott értékek közötti szoros korreláció, mely megerősíti a szignifikáns szavak alapján történő kivonatolás létjogosultságát.

Jegyzetek

1. KOLTAY Tibor: Szöveg, információ, relevancia: néhány adalék a témakörhöz. In: Könyvtári Figyelő. 2005. (51. évf.) 3. sz. pp.514–518.
2. PROKNÉ Palik Mária: A tartalmi feltárás problémái online könyvtári katalógusokban. In: Tudományos és Műszaki Tá-
3. Szeretném megköszönni a Szegedi Egyetemről Barátné Hajdu Ágnes, valamint a Debreceni Egyetemről Bényei Miklós és Boda István segítségét.
4. A felmérés időpontjában a konzultáción jelen lévő hallgatók.
5. A felmérés elérhető a www.ekt.hu/1/1mt/ webcímen.
6. A felmérés során megjelölt releváns szavak részletes elemzése a következő kiadványban jelenik meg: Lengyelne Molnár Tünde: Az emberi és gépi referátumkészítés hatékonyságának elemzése. In: Agria Media 2006. Konferencia Kiadvány. Eger: Líceum Kiadó (megjelenés alatt)
7. Ha elhagyjuk azokat a hallgatókat, akik nem jelöltek meg egyetlen mondatot sem. Az érték azon hallgatók figyelembevételével, akik a következő cikknél elkészítették a kivonatot, de a Koltay-cikk kivonatánál nem jelöltek meg egyetlen mondatot sem: 16 és 28
8. A mondatok sorrendjét alapul véve.
9. Koltay Tibor cikke esetén:0,769775; Prokné Palik Mária cikke esetén: 0,769793
10. Informatikus-könyvtáros főiskolás hallgatóknál 2 mondat, egyetemista mintacsoportnál 5 mondat, szakembereknél 1 mondat.
11. 59,52%
12. Azonos számú jelöléssel.



Az Országos Széchényi Könyvtár gondozásában várhatóan 2007 novemberében megjelenik a *Régi Magyarországi Szerzők (RMSZ) I. A kezdetektől 1700-ig c.* kiadvány.

Az RMSZ első változata 1989-ben jelent meg kéziratban Szabó Géza adatgyűjtésének felhasználásával Pintér Gábor szerkesztésében, elsősorban az OSZK-ban folyó bibliográfiai gyűjtőmunka segédeszközeként. A kutatók és bibliográfusok részéről hamarosan megfogalmazódott az igény, hogy az adatokat pontosítani és bővíteni kellene.

Jelen kötet azoknak a magyarországi

szerzőknek az életrajzi adatait tartalmazza betűrendben, akik a kezdetektől 1700-ig éltek, ill. a 18. században haltak meg, de 1701 előtt már ismeretes legalább egy nyomtatott munkájuk. A korábbi kötetben 3782 magyarországi szerző adata szerepelt, a mostaniban 6484 szerző adatai találhatóak, 8076 rájuk vonatkozó utalóval. A munkálatokat Wix Györgyné kezdte el, s az RMNY Szerkesztőség munkatársai Kis Elemérné, Kovács Zsuzsanna és P. Vársárhelyi Judit fejezték be.

További információk a kiadvány megrendelhetőségéről: kiadvany@oszk.hu

(Katalist, 2007. június 8.)