

A magyar igazságügyi könyvszakértői vélemények bírósági befogadásának vizsgálata gépi tanuló algoritmussal

DOI: [10.29180/978-615-6342-50-8_5](https://doi.org/10.29180/978-615-6342-50-8_5)

SZERZŐK:

Erdélyi Katalin¹, Fauszt Tibor², Kovács Endre³, Dobák Dóra⁴

ABSZTRAKT

A 2000-es években indult Big Data forradalom óta folyamatosan sokszorozódik a digitális tartalom. A jelenség kutatók tízezreit sarkallja szerte a világon a nagy mennyiségű adatból kinyerhető új és hasznos információk megszerzésére. A tömörített digitalizált dokumentum szövegében rejlő rejtett összefüggések feltárására matematikai, statisztikai, valószínűségszámítási és számítástudományi algoritmusok léteznek, vagy lehet akár egyedi algoritmust is fejleszteni. Adatbányászati módszerekkel viszonylag egyszerű az olyan dokumentumok szövegtörzsének elemzése, amely azonos vázzal, közel szabványos módon, formanyomtatvány jelleggel rendelkezik. Ilyen iratok például az oktatásban is használt jegyzőkönyvek vagy a bíróságok által hozott határozatok. Az ezekhez hasonló, félig strukturált szövegek automatikus feldolgozására kezdtünk el egy adatbányászati alapokon nyugvó saját algoritmust fejleszteni, amelyet később önálló döntéshozásra szeretnénk „rábírní”. Módszerünket a bírósági határozatokra, mint szövegtörzsekre munkáltuk ki, azok közül is azokon szeretnénk a tesztelést elvégezni, amelyek könyvszakértőt említenek.

Kulcsszavak: szövegbányászat, bírósági ítélet, könyvszakértő, digitális dokumentum

ABSTRACT

The Big Data revolution that began in the 2000s has continuously multiplied the amount of digital content. The phenomenon drives thousands of researchers worldwide to acquire new and valuable information from large amounts of data. There are mathematical, statistical, probability and computer science algorithms to reveal hidden connections in the text of the vast number of digitized documents, or even a unique algorithm can be developed. Data mining methods make it relatively easy to analyze the text corpus of documents with the same structure, are nearly standard and are in the form of forms. Such documents, for example, are also used in education, minutes or judgments passed by courts. We have started developing our algorithm based on data mining for the automatic processing of semi-structured texts similar to these, which we later want to "persuade" for independent decision-making.

¹ Budapesti Gazdasági Egyetem, Pénzügyi és Számviteli Kar,
ORCID: 0000-0001-5010-6688, e-mail: erdelyi.katalin@uni-bge.hu

² Budapesti Gazdasági Egyetem, Pénzügyi és Számviteli Kar, Gazdaságinformatica Tanszék,
ORCID: 0000-0003-3590-1100, e-mail: fauszt.tibor@uni-bge.hu

³ Budapesti Gazdasági Egyetem, Pénzügyi és Számviteli Kar, Gazdaságinformatica Tanszék,
ORCID: 0000-0001-6030-0716, e-mail: kovacs.endre@uni-bge.hu

⁴ Budapesti Gazdasági Egyetem, Pénzügyi és Számviteli Kar, Gazdaságinformatica Tanszék,
ORCID: 0000-0002-0857-9999, e-mail: dobak.dora@uni-bge.hu

We have developed our method for court judgments as text corpora, and we want to carry out the testing which mention audit experts.

Keywords: text mining, court judgment, audit expert, digital document

Az igazságügyi könyvszakértői tevékenység

A büntetőeljárás törvény szerint az igazságszolgáltatás a bíróságok feladata. (Be. 11. §., 2017). A bíróság ítélete csak akkor lehet igazságos, ha lehetősége van az események valósághű megismerésére, ebben különféle bizonyítási eszközök - okirati bizonyítékok, tanúvallomások szakértői vélemények - segítik. (Erdélyi, 2022) Gazdasági bűncselekmények esetén, mint például alaptőke-emelési joggal való visszaélés (Kúria, 2021), vagy a vállalkozás fizetéseképtelensége esetén a valós pénzügyi helyzet vizsgálata (Fővárosi Törvényszék, 2017), vagy akár szállítási tevékenység miatt (baleset) elmaradt haszon értékének megállapítása végett (Jász-Nagykun-Szolnok Megyei Bíróság, 2008) a bíróság igazságügyi könyvszakértői véleményt szerezhet be. Igazságügyi könyvszakértők nyilatkoznak minden adó-, illeték-, ár-, pénzügyi, banki és könyvviteli szakértői kérdésben (Fodor, 2014.); egyébként pedig a könyvvizsgálat a piacgazdaság egyik jelentős szereplője a szervezetek általános ellenőrzési tevékenysége által (Lukács & Török, 2019).

A magyar igazságügyi szakértői tevékenység minőségéről általában véve nincsenek publikus adatok, feltehetően nem is kerül mérésre. „Az Igazságügyi Minisztérium ugyan gyűjti az igazságügyi szakértők ügyforgalmára vonatkozó statisztikai adatokat (IM, 2016) , ám ezeket nem teszi közzé.” (Erdélyi, 2022, old.: 106) Az nyilvánvaló, hogy a szakértők különleges szakismereteik révén segítik a bűncselekmények felderítését és a tényállás bizonyítását (Kovács, 2020a.), de joggal merül fel a kérdés: milyen mértékben? Szavuk szent lehet, így szakvéleményük egyértelműen és minden esetben az ítélezés alapját képezik, vagy időnként valamilyen oknál fogva véleményük kizárásra kerül a bizonyításból? (Kovács & Nogel, 2017) Hogy milyen mértékű lehet az elfogadott/elutasított szakvélemények aránya, erre vonatkozóan sincsenek mért és nyilvánosságra hozott adatok. Annyit tudunk, hogy „az elmúlt két évtizedben tudományos jelentések sora hívta fel a figyelmet a forenzikus módszerek gyenge pontjaira, a téves szakvéleményekre alapozott téves ítéletekre, és a megoldásra váró problémákra.” (Kovács, 2021., old.: 1741)

A magyar bíróságok ítéletadatbázisa

Egy kutatót mindig izgalommal tölti el az, ha sehogy sem talál olyan adatot, ami egyébként kiszámítható. Szakértői területen sok nem létező, de egyre inkább kiszámítható adat lehet, álláspontunk szerint egyike a sokaknak az elfogadott/elutasított szakértői vélemények aránya. Az ehhez szükséges adatforrás a magyar Bírósági Határozatok Gyűjteménye, amely tárolja és mindenki számára elérhetővé teszi a bíróságok által hozott határozatokat. Ezekből a dokumentumokból nemcsak a magyar bírósági ítélezés sajátosságai rajzolódhatnak ki, hanem a jogalkalmazók szakértői teljesítményéhez való viszonyulása, elégedettsége is, mégpedig valósághűen, hiszen az ítéletek jó része tartalmaz valamilyen szakértői véleményre történő hivatkozást, jó esetben szakértői vélemény értékelést. Az ítéletadatbázis létrehozásának egyik célja az volt, hogy a magyar bíróságok működésének átláthatóságát biztosítsa, elsősorban a lakosság, illetve a jogalkalmazók számára (Eötvös Károly Intézet, 2009), a bíróságoknak

pedig törvényi kötelezettségük a folyamatosan keletkező ítéletek feltöltése (XC. tv., 2005). Az adatbázis 1996-tól napjainkig tartalmazza a magyar bíróságokon keletkező ítéletek anonimizált változatát, a határozatok leválogatását pedig a bhgy.hu webes felület szokványos keresőmotorja biztosítja. Az ítéletek szűrésére megadhatók kulcsszavak, időintervallumok, és olyan konkrétumok alapján is lehet keresni, mint a határozatot meghozó bíróság neve, az érintett jogterület (büntető, polgári, gazdasági jog stb.), vagy az ítélezési gyakorlatra történő hivatkozás.

Bírósági ítéletadatbázisok más országokban is léteznek és kutatottak (Li & Li, 2019). Az igazságügyi szakértői véleményre, mint bizonyítási eszközre vonatkozó nemzetközi joganyag is megismerhető, így a véleményekre vonatkozó ítéleti adatok nemzetközi viszonylatban is összehasonlíthatók lehetnek. (Kovács & Nogel, 2017) A magyar BHGY is kedvelt adatforrás, Zódi Zsolt 2017-ben publikált arról, hogyan találtak a BHGY ítéleteiben korábbi esetekre történő hivatkozási mintázatokat, azt megtudni, „...hogyan nem mozognak-e a kontinentális jogrendszerek, és benne a magyar jogrendszer a precedensjog irányába?” (Zódi, 2017, old.: 4). Egy idei cikk pedig a bűnügyi célú DNS minták szennyezettségi mértékének vizsgálatát írja le (Kovács, Erdélyi, Pádár, Zenke, & Nogel, 2022), ugyancsak a BHGY dokumentumainak elemzése alapján. Zódi cikkének idején a BHGY 60 ezer határozatot tartalmazott, ez mára több mint kettőszázezerre nőtt, azaz 5 év alatt a BHGY-ba feltöltött dokumentumok száma megháromszorozódott.

Az adatbázisban az ítéletek számítógépes szöveges fájlok, Rich Text dokumentum-fájlformátummal, amelyek letöltési lehetőséggel rendelkeznek. Terjedelmüket tekintve átlagosan körülbelül 15.000 szóból állnak, felépítésük szerint pedig félig strukturáltak.

1. ábra Egy BHGY-ből letöltött ítélet, jellemzően félig strukturált felépítéssel

Fővárosi Törvényszék 29.G.40.266/2017/102.
A Fővárosi Törvényszék
a Ruszthy Ügyvédi Iroda (cím - ügyintéző: dr. Ruszthy Gyula ügyvéd) által képviselt felperes neve (felperes címe) felperesnek
a Holczer, Jákó & Boross Ügyvédi Iroda (cím - ügyintéző: dr. Boross Ildikó ügyvéd) által képviselet alperes neve (alperes címe.) alperes ellen
ügyvezető, végelszámoló felelősség megállapítása iránt indított perében meghozta az alábbi
Í T É L E T E T:
Az alperes a B. K. F. Korlátolt Felelősségű társaság vezető tisztségviselőjeként a társaság fizetési képzetlenséggel fenyegető helyzete bekövetkezését követően ügyvezetési feladatait 2007. február 20. napjáig a magáncélú kiadások körében nem a hitelezői érdekek elsődlegessége alapján látta el, s ezáltal a társaság vagyona 415.723,- (négy százötvenöt ezernégy százhatvanhárom) forinttal csökkent.
A bíróság kötelezi az alperest, hogy 15 napon belül fizessen be a B. K. F. Korlátolt Felelősségű társaság „felszámolás alatt” cég vagyonához tőke hozzájárulás címén 80.051.664,- (nyolcvanmillió-ötvenegyezer-hatszázhatvanégy) Ft-ot a O. G. Sz. T. Kft. felszámoló Bank Nyrt.-nél vezetett ...letéti számlájára történő befizetés útján.

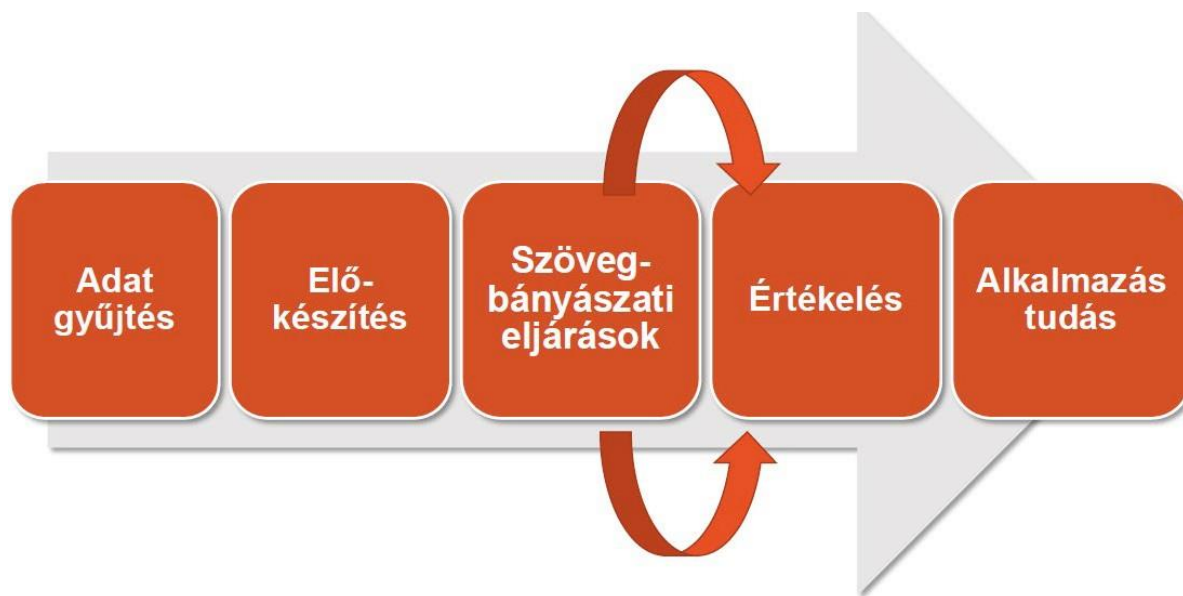
Forrás: Bírósági Határozatok Gyűjteménye

Szövegbányászat

A szövegbányászat szöveges adatokon végzett feldolgozási és elemzési tevékenység új információk, mintázatok, rejtett összefüggések feltárása, azonosítása és elemzése végett (Tikk, és mtsai., 2007). A szövegbányászatban maga a szöveg az adat, a strukturált szöveges adathalmaz pedig a korpusz. A szövegek bányászatának nehézségét azt jelenti, hogy az emberi kommunikáció nem a számítógépes feldolgozás szerint alakult ki. Az emberi elme a nyelvi mintázatok felismerésében jeleskedik, míg a gép a nagy mennyiségű adat gyors feldolgozásában. A kettő kombinációja, az emberi elme és a gép kapacitása által küszöbölődnek ki a felek hiányosságai, feltéve, hogy sikerül a gépnek elmagyarázni az emberi szándékot. Jó esetben az automatizált szövegelemzés jelentős mértékben képes csökkenteni az elemzéshez szükséges időt és emberi munkát.

Ami a szövegbányászat módszereit illeti, az egyszerű statisztika nem elegendő; mint ahogyan Zódi értelmezésében is történt, a szöveges adatbázisok anyagára Big Data-ként kell tekinteni, aminek feldolgozásához adatbányászati technológiák szükségesek, főként a gépi tanulás eszköztára, algoritmikus megoldásokkal. Viszont megéri: az eredményt olyan trendek, mintázatok, összefüggések jelentik, különösen az emberi viselkedésről, interakciókról, amelyek más módszerekkel rejtve maradnának. (NIST Big Data Public Working Group, 2015). A szövegbányászat általános modelljét mutatja az alábbi ábra (Pancza, 2019).

2. ábra A szövegbányászat általános modellje



Forrás: Pancza Judit: Szövegbányászat – SPSS Nyári Iskola (2019)

Amikor egy szövegbányászati feladatban pontosan ismertek és ellenőrizhetők az elvárt eredmények, a szöveg automatikus feldolgozásához felügyelt gépi tanulás alkalmazható. Ennek célja egy olyan modell megalkotása és betanítása, ami korábban még nem látott esetekre is helyesen működik. A modellépítés annak a döntési szabálynak a megalkotását jelenti, ami az egyedek tulajdonságai között feltárt szabályszerűségeket tartalmazza. A modell fogja a korábban még nem látott egyedekre a döntést meghozni. A tanulási folyamatot Sebők és társai a következőképpen írják le: „A felügyelt tanulási feladat megoldása egy úgynevezett tanító halmaz (training set) meghatározásával kezdődik, melynek során a kutatók saját maguk végzik el kézzel azt a feladatot, melyet a továbbiakban gépi

közreműködéssel szeretnének nagyobb nagyságrendben, de egyben érvényesen (validity) és megbízhatóan (reliability) kivitelezni. Eredményeinket az ugyanúgy eredetileg kézzel lekódolt, de a modellépítés során félretett teszhalmazunkon (test set) értékelhetjük.” (Sebők, Ring, & Máté Ákos, 2021, old.: 159).

A felügyelt tanulás egyik tipikus feladattípusa az osztályozás, amely az egyedeket az előre megadott osztályok valamelyikébe sorolja, és ellátja az osztálycímkével. A gépi tanuló algoritmusok a nyers adatokból nem közvetlenül tanulnak, szükségük van a vizsgált jelenségek numerikus jellemzésére. Így definiálni kell az egyedek releváns információt tartalmazó attribútumait, hogy a gépi tanuló modell az attribútumok közti összefüggéseket, szabályszerűségeket megtanulhassa.

Hogy vajon teljesül-e a „több tapasztalattól jobb teljesítmény” gépi tanulási alapelv, azt az osztályozási feladatmegoldás jóságát mérő találati arány, mint kiértékelési metrika fogja megmutatni. (Bognár & Fauszt, 2022) Ez annak mérését jelenti, hogy hányszor sorolta a modell a helyes vagy helytelen osztályba az új egyedeket.

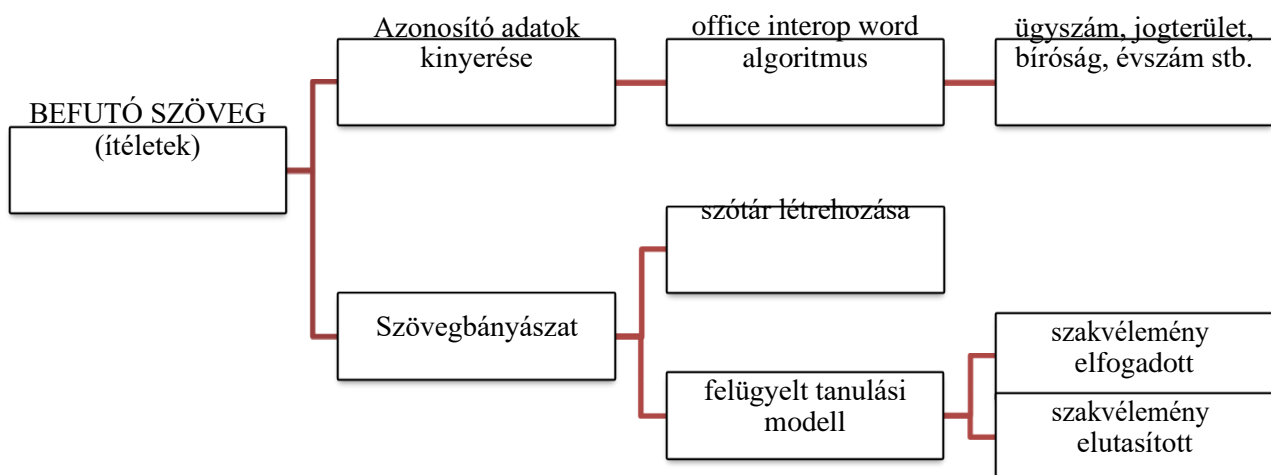
A szakértői vélemények minőségi kérdéseinek kutatása

Célok és módszerek

A szakértői vélemények megfelelősége, minősége, validitása komplex kérdés (Kovács, 2017). Az egymástól jelentősen különböző fogalmak részletes elkülönítését és definícióját Nogel Mónika monográfiájában megtaláljuk (Nogel, 2020).

Jelen tanulmányunk azért készült, hogy a szakértői vélemények minőségi kérdéseinek kutatásával összefüggő vizsgálataink jelenlegi állapotát bemutassa. Általánosságban pedig az a célunk, hogy a magyar bíróságok által hozott ítéletekből kinyerjük a jogalkalmazók szakértők munkájával kapcsolatos nézeteit, a későbbiekben pedig feltárjuk az elutasított szakértői vélemények okait. Vizsgálataink egyik állomása az, hogy megállapítjuk az elfogadott/elutasított szakvélemények arányát. Ez egy osztályozási feladat, hiszen az ítéleteket két halmazba kell különítenünk, egyikbe azok kerülnek, amelyekben a könyvszakértő véleményét nem fogadták el, a másikba azok, amelyekben elfogadták.

3. ábra Célok és módszerek



Forrás: Saját szerkesztés

Tekintve az elemzendő szöveges minta Big Data jellegét, tanuló algoritmust fejlesztünk. Ehhez a Microsoft.Office.Interop.Word algoritmusát, szótáralapú keresést és felügyelt tanulási modellt használunk (1. ábra).

A felhasznált adatok

Kutatásunk számunkra releváns dokumentumai a BHGY gazdasági büncselekmények bizonyításáról szóló határozatai, ezek közül is azok, amelyekben könyvszakértőre való hivatkozás történik. A könyvszakértőre hivatkozó bírósági ítéletek összessége a teljes szövegtörzsünk, az egyes ítéletek, azaz az egyes dokumentumok pedig ennek részegységei. Elemzésünk tématerülete a könyvszakértői vélemények elfogadásra/elutasításra vonatkozó nyelvi elemek, ezek határozzák meg a szöveg szókészletével szembeni elvárásainkat. A szóhasználat felderítésére szótáralapú összehasonlítást végzünk.

A gazdasági perek mennyiségi adatairól azt lehet tudni, hogy számuk Magyarországon 2000-2020 között körülbelül 22.000 volt⁵, aminek körülbelül a 8%-a, mintegy 1800 ítélet említi a könyvszakértő tárgyalásba való bevonását. A BHGY adatbázis felületéről, többszörös szűkítéssel kaptuk meg ezeket az adatokat, és töltöttük le szisztematikusan az 1800 ítéletet.

A kutatás lefolytatása - adatkinyerés

Szövegtörzsünket az 1800 darab bírósági ítélet alkotja, ezen az adatbázison végzünk szövegbányászatot, gépi tanuló megoldással. A szövegtörzsből, mint nyers adatokból felügyelt tanulási modellt szeretnénk létrehozni. A modellépítés első lépése a könyvszakértői vélemények elfogadásával/elutasításával összefüggő adatok (jellemzők, információk – úgynevezett attribútumok) szövegtörzsből való kinyerése implementálással (gépi kódolással). Ez a jellemzőkészlet képezi a gépi tanuló rendszerünk egyik felét.

A kinyert adatok a következők:

- a határozat sorszáma (pl. 29.G.40.266/2017/102.)
- a határozat által érintett jogterület (pl. gazdasági jog)
- a határozatot meghozó bíróság neve (pl. Fővárosi Törvényszék)
- a határozat keletkezésének évszáma (pl. 2020)
- a szakértői terület típusa (pl. könyvszakértés)
- a határozat elvi tartalma, amennyiben a dokumentumban szerepel.

A felsorolt adatokat a .NET keretrendszer által biztosított Microsoft.Office.Interop.Word (OIW) programcsomag eljárásaival vonjuk ki a szövegekből, az eljárások a szövegek fájljait hierarchikus mapparendszerben tárolt saját számítógépről érik el. Az OIW osztálygyűjtemény text típusú fájlok manipulációját lehetővé tevő függvényeket tartalmaz, azokat, amelyeket a Microsoft Word szövegszerkesztője is használ. Az OIW biztosítja a megszokott Word műveleteket, mint például bekezdések és szavak megkeresését, másolást, kivágást, beillesztést. A használt programnyelvek a C# és az SQL.

⁵ Forrás: Bírósági Határozatok Gyűjteménye (bhgy.hu)

A kutatás lefolytatása – tanuló modell építés, jellemzőkinyerés

A gépi tanulórendszerünk másik felét alkotó jellemzőkészlet szövegtörzsből való kinyerésére egy saját magunk által összeállított szakmai szótárt veszünk igénybe. A szövegbányászati elemzés gyakran használt eleme a szótár. A szótár olyan szószerkezetet jelent, amely szavai segítik a kutatás szempontjából releváns tartalmak azonosítását. Léteznek előre elkészített szótárak, de ha a tématerület specifikus nyelvezettel rendelkezik, mint az jelen esetben történik (jogi nyelvezet), a kutatóknak saját magának szükséges azt létrehozni az elemzendő szövegtörzsből. Ez igen fáradtságos munka. A mi szótárunkba az ítéletek szövegéből kiemelt, a könyvszakértői vélemények elfogadására/elutasítására vonatkozó szavak, szóösszetételek, több szóból álló kifejezések vagy félmondatok kerülnek. Először gépi úton, az OIW függvényeivel kinyerjük az ítéletek „könyvszakértő” szót tartalmazó bekezdéseit, az eredeti szövegtörzshöz képest egy kisebb részhalmazát képezve. Ebből a továbbiakban hagyományos módon válogatjuk ki a szótárelemeket, azaz azokat a kifejezéseket, amelyeket nyelvileg a szakértői vélemény elfogadására vagy elutasítására utalónak ítélnék. Szótárunk alapját a következő táblázatban a teljesség igénye nélkül bemutatott mondatrészek alkotják. A táblázat mondataiból jól érzékelhető a szakvélemények elfogadására/elutasítására vonatkozó megfogalmazások nyelvi változatossága. Természetesen a legkönnyebb dolgunk akkor lenne, ha az ítéletek egyértelműen tartalmaznák azt, hogy a „szakvélemény elfogadott” vagy a „szakvélemény elutasított”, de ez csak ritkán van így. A megfogalmazás nyelvi variálódása miatt szükséges a szótár létrehozása és az ítéletek szövegének szótáron való megfuttatása.

1. táblázat Szótárunk alapját képező mondatrészek

Az ítéletek változatos megfogalmazásai a szakvélemények elfogadására/elutasítására

... a könyvszakértő kellően alátámasztott, részletesen megindokolt, aggálytalan szakvéleménye alapján a bíróság arra a következtetésre jutott...

... a bíróság a szakértői vélemények egybehangzósága és kategorikussága okán elfogadta...

... a könyvszakértői vélemény nem egyértelmű, kellően nem megalapozott, így bűnösség megállapítása alapjául nem szolgálhat...

... a bíróság a könyvszakértői véleményt az ítélezés alapjául elfogadta...

... a könyvszakértői vélemény megállapításai és az ezeket vitató alperesi állítások a perben relevanciával nem bírnak...

... a könyvszakértő kiegészített szakvéleményét a bíróság a bizonyítékok köréből mellőzte...

... a bíróság figyelmen kívül hagyta a könyvszakértői véleményt...

... a könyvszakértői vélemény ellentmondásait a szakértőnek nem sikerült feloldani...

... bíróság elfogadta a könyvszakértő személyes meghallgatása során kifejtett álláspontját...

... a perben beszerzett könyvszakértői vélemény alátámasztotta ...

Forrás: saját gyűjtés, saját szerkesztés

A jellemzőkinyerést követően meghatározzuk a tanító adatbázis döntési szabálykészletét, és kézzel felcímkézzük a konkrét egyedeket leíró úgynevezett jellemzővektorokat, azaz attribútumpárosokat. Esetünkben az egyed az egy bírósági ítélethez tartozó összes attribútum (2. sz. táblázat egy sora), az attribútumok pedig a táblázat fejlécében található ítéletjellemzők (sorszám, jogterület, bíróság, év, prediktor stb.). A folyamat eredményeképpen logikailag az alábbi táblázat áll össze. A teljes szövegkorpuszunk mintegy 20-30%-a fogja kitenni az ezen az elven szerveződő tanító adatbázist.

2. táblázat Tanítóhalmazunk egy részlete

Határozat sorszáma	A hározat elvi tartalma	Jogterület	Bíróság	Év	Szakértő	Bekezd	Mondat	Prediktor	Osztálycímke
Gfv.30407/2021/4	Deviza alapú kölcsönszerződéssel kapcsolatos tisztességtelen tájékoztatás.	gazdasági	Kúria	2022	könyv	1	A bíróság a könyvszakértő kizárását külön végzésben indokolva megtagadta.	-	-
Gfv.30407/2021/4	Deviza alapú kölcsönszerződéssel kapcsolatos tisztességtelen tájékoztatás.	gazdasági	Kúria	2022	könyv	3	A bíróság megállapította, az alperes a pusztán előadásával nem tudta kétségessé tenni a könyvszakértő számításait.	nem tudta kétségessé tenni a könyvszakértő számításait	Elfogadott
Gfv.30407/2021/4	Deviza alapú kölcsönszerződéssel kapcsolatos tisztességtelen tájékoztatás.	gazdasági	Kúria	2022	könyv	4	A bíróság a könyvszakértői véleményt az ítélezés alapjául elfogadta.	könyvszakértői véleményt elfogadta	Elfogadott
Gfv.30407/2021/4	Deviza alapú kölcsönszerződéssel kapcsolatos tisztességtelen tájékoztatás.	gazdasági	Kúria	2022	könyv	5	A bíróság az igazságügyi könyvszakértő álláspontját elfogadta.	könyvszakértő álláspontját elfogadta	Elfogadott

Forrás: saját szerkesztés

A tanító adatbázis összeállítását követően következhet a tanuló adathalmaz betanítása. A tanítás eredménye az lesz, hogy a még fel nem dolgozott, ismeretlen egyedek emberi beavatkozás nélkül, automatikusan megkapják az „Elfogadott” vagy „Elutasított” osztálycímkeket.

A kutatás lefolytatása – a modell validálása

A felépített tanuló modell teljesítményének mérése mindig kulcsfontosságú a gépi tanulási feladatokban. Ehhez a validálási tevékenységhez az osztályozás jóságát mérő találati arányt, mint kiértékelési metrikát fogjuk használni, azaz azt fogjuk mérni, hányszor sorolta a modell helyesen vagy helytelenül valamelyik osztályba az új egyedeket. Az osztályokba sorolás lehetséges kimeneteit a 3. táblázat mutatja.

Rendszerünk gyenge teljesítménye esetén megkeressük az ellentmondásokat, hibákat és azok okait, hogy a modellt javítani tudjunk. A javítás inkrementálisan fog történni, és minden javítás után újra teszteljük az osztályozás pontosságát. Ha egy rendszer rossz összefüggéseket tanul meg a tesztadatsoron, az éles adatokon kiszámíthatatlan pontatlanságot fog eredményezni. A pontosság javításának egyik lehetséges módja a szótár bővítése, módosítása, illetve a tanításra szánt adathalmaz túltanítás határáig való növelése.

3. táblázat Az ítéletek osztályokba sorolásának lehetséges eredményei

A könyvszakértői vélemények elfogadására/elutasítására alkotott modell lehetséges kimenetei		
Találat	Modell általi osztályozás	Valóság
<i>True pozitív</i>	Elfogadott	Elfogadott
<i>Fals pozitív</i>	Elfogadott	Elutasított
<i>True negatív</i>	Elutasított	Elutasított
<i>False negatív</i>	Elutasított	Elfogadott

Forrás: saját szerkesztés

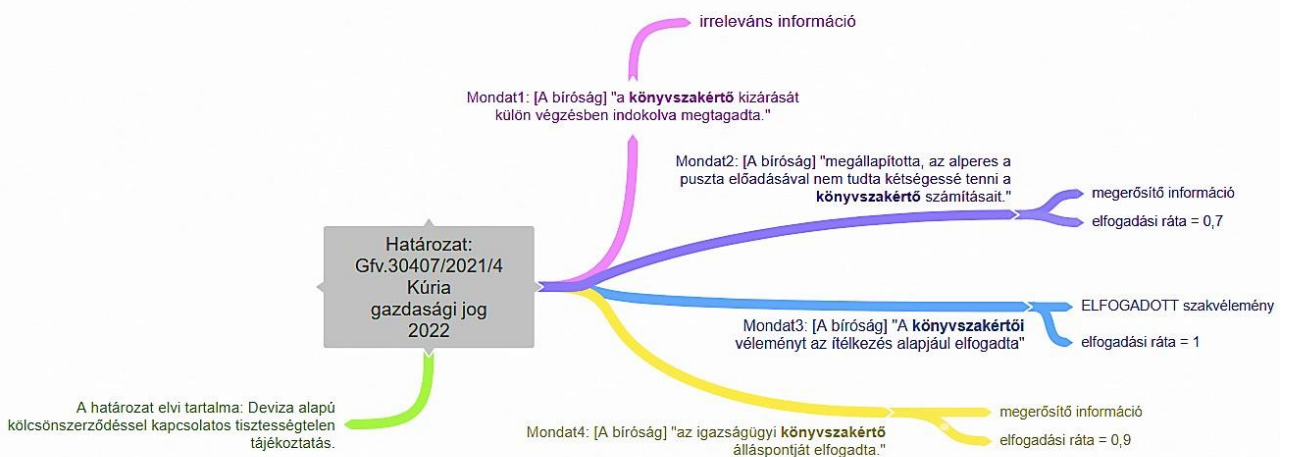
A kutatás kihívásai

A ráutaló kifejezések ráutalásának mértéke

A szótár bőségének az osztályozás pontosságában nagy szerepe van, de sajnos a szótárban való keresés és találat sem mindig elegendő a tanító adathalmaz létrehozásához. Esetünkben a felcímkezéshez figyelembe kell venni az azonos bekezdésben lévő ráutaló szavak egymásutánosságát, vagy egymástól való távolságát is. Másrészt pedig a szótár szavai sem egyforma mértékben utalnak a szakértői vélemények elfogadására/elutasításra, ezért valószínűleg valamilyen súlyozási megoldást is be kell vezetnünk, egy „elfogadási ráta”-szerűséget, ami a címkére való utalás mértékét jelöli.

Az alábbi ábrák két bírósági ítélet könyvszakértőt említő bekezdéseinek logikai értékelését mutatják.

4. ábra Döntési logika – 4 mondat, egyszerű eset



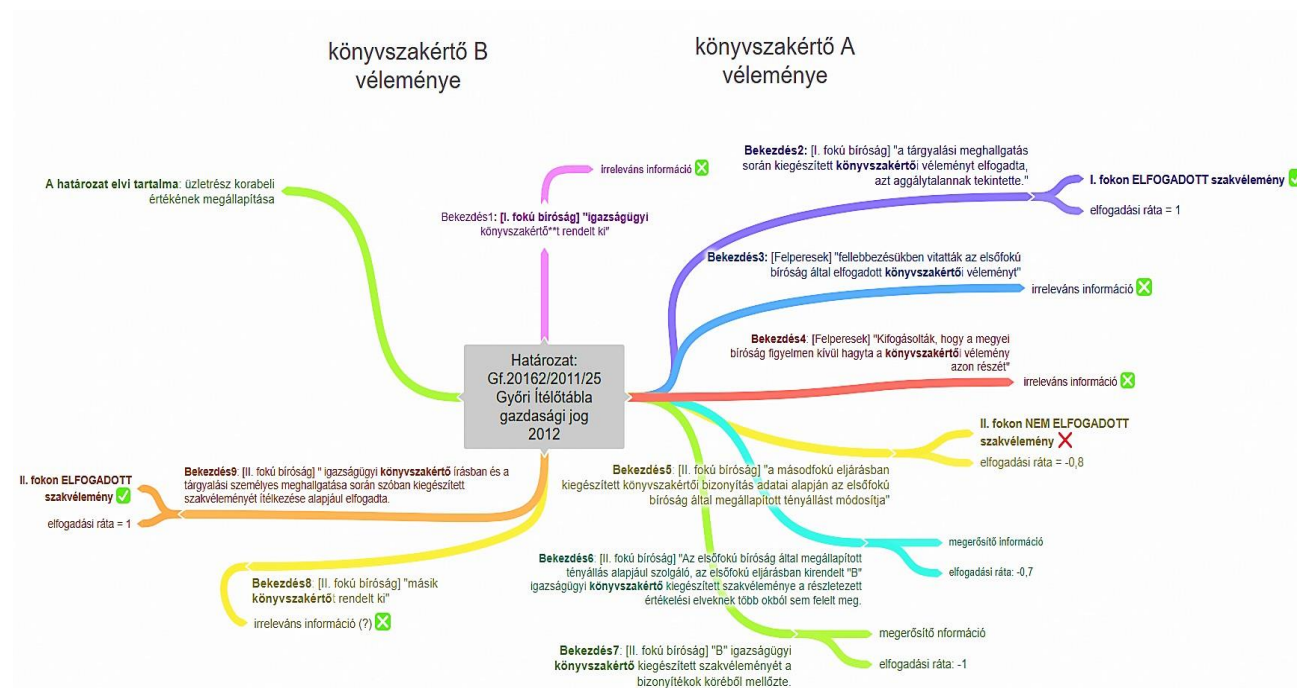
Forrás: Saját szerkesztés

A 4. sz. ábrán látható határozat mindössze 4 mondata szól könyvszakértőről, közülük az elsőnek irreleváns az információ tartalma a szakvélemény elfogadására vonatkozóan, a 2. mondat pedig arra utal, hogy az ominózus szakvéleményt elfogadta a bíróság, de a szöveg nyelviileg pontosan nem említi

az „elfogadás” szót. A 3. mondatban egyértelműen szerepel az „elfogadta” szó, a negyedik pedig kimondás nélkül erősíti meg az elfogadást. A szakvélemény elfogadottságát a 3. mondat egyértelműsíti, a többi megerősíti, és semmilyen ellentmondás nincs.

Az 5. sz. ábra egy bonyolultabb esetről szól: a vizsgált ítéletben összesen 9 bekezdésben említett könyvszakértő két különböző könyvszakértő (könyvszakértőA és könyvszakértőB), két különböző tárgyalási szinten, hiszen elsőfokú és másodfokú bíróság is tárgyalta és vett igénybe az ügyben könyvszakértőt, amire az ítélet hivatkozik is. Az egyik szakértő véleményét elfogadták, de a másikat nem – ezt kell az algoritmusunknak az ítélet szövegéből felismernie.

5. ábra Döntési logika – 9 bekezdés, bonyolult eset



Forrás: Saját szerkesztés

Összefoglalás

Tanulmányunk által egy folyó kutatás kulisszatitkaiba pillanthatott bele az Olvasó. A kutatás elején járunk, a problémát felismertük, az adatokat összegyűjtöttük, a céljaink egyértelműen körvonalazódtak, és némely részfeladatra már kész megoldásaink vannak. Abban biztosak vagyunk, hogy nem triviális a feladat, de amennyiben modellünk sikeresnek bizonyul, más szakértői területeken is alkalmazni lehet. Eredményeink jelentőségét az adja, hogy a Bírósági Határozatok Gyűjteményének automatikus szövegelemzése által megismerhető a peres ügyek lefolyása, a bírói döntést alakító számos tényező, a szakértők bírósági perekben betöltött valós szerepe és szakvéleményük hiteltérdemlősége.

Irodalomjegyzék

1. Be. 11. §. (2017).
2. Bognár, L., & Fauszt, T. (2022). Factors and conditions that affect the goodness of machine learning models for predicting the success of learning. *Computers and Education: Artificial Intelligence* 3 Paper: 100100.
3. Eötvös Károly Intézet. (2009). Az igazságszolgáltatás nyilvánossága különös tekintettel a bírósági határozatok nyilvánosságára. Letöltés dátuma: 2022.. 10. 21., forrás: http://ekint.org/lib/documents/1479484404-a_targyalasok_nyilvanossaga.pdf
4. Erdélyi, K. (2022). Az írásszakértői vélemények relevanciája Magyarországon. *Jog-Állam-Politika - Jog- és politikatudományi folyóirat*, 2022/3.
5. Farkas, R. (n. i.). Gépi tanulás alapfogalmai. Letöltés dátuma: 2022.. 11. 22., forrás: Gépi tanulás a gyakorlatban: <https://www.inf.u-szeged.hu/~rfarkas/ML20/alapfogalmak.html>
6. Fodor, R. (2014.). Az igazságügyi könyvszakértés jelentősége napjainkban. Letöltés dátuma: 2022. november 20., forrás: ujbtk.hu: <https://ujbtk.hu/fodor-reka-az-igazsagugyi-konyvszakertes-jelentosege-napjainkban/>
7. Fővárosi Törvényszék. (2017). 29.G.40.266/2017/102. sz. ítélet. Forrás: Bírósági Határozatok Gyűjteménye.
8. IM. (2016). Az igazságügyi szakértők statisztikai adatszolgáltatásáról, a kamarai költségátalány bevallásának és megfizetésének szabályairól, valamint az igazságügyi szakértő által vezetett adattovábbítási nyilvántartásról szóló 13/2016. (VI. 29.) IM rendelet.
9. Jász-Nagykun-Szolnok Megyei Bíróság. (2008). 11.G.20.171/2008/70. sz. ítélet. Forrás: Bírósági Határozatok Gyűjteménye.
10. Kovács, G. (2017). Az Európai Forenzikus Tudományos Térség (EFSA-2020) megalkotásának koncepciója. *Jog-Állam-Politika*, 2017/1, 83-101.
11. Kovács, G. (2020a.). Igazságügyi szakértő alkalmazása a büntetőeljárásban. In S. Zsuzsa (Szerk.), *Szakkérdésekről szakértőknek – Igazságügyi szakértőkre vonatkozó jogi kérdések tudományos és gyakorlati megközelítésből*. Budapest: Magyar Közlöny Lap- és Könyvkiadó.
12. Kovács, G. (2021.). Forenzikus tudományok ostrom alatt. *Belügyi Szemle*, 69(10), 1741-1758. doi:<https://doi.org/10.38146/BSZ.2021.10.4>
13. Kovács, G., & Nogel, M. (2017). Activity of Forensic Experts and Quality Assurance in Hungary. *Forensic Science International*.
14. Kovács, G., & Nogel, M. (2017). Comparative Analysis of the Legal regulation of Forensic Experts in Europe. *Forensic Science International* 277, 81-82.
15. Kovács, G., Erdélyi, K., Pádár, Z., Zenke, P., & Nogel, M. (2022). Contamination of forensic DNA evidence in the light of Hungarian court. *Forensic Science International: Genetics Supplement Series*.
16. Kúria. (2021). Gfv.VI.30.407/2021/4. sz. ítélet. Forrás: Bírósági Határozatok Gyűjteménye.
17. Li, B., & Li, N. (2019). Handwriting expertise reliability: A review. *Journal of*, 5(4).
18. Lukács, L. I., & Török, M. Z. (2019). A magyar könyvvizsgálat történetének mérföldkövei a középkortól napjainkig. *Polgári Szemle*, 212-223. doi:DOI: 10.24307/psz.2019.1214
19. NIST Big Data Public Working Group. (2015. September). NIST Big Data Interoperability Framework: Volume 1, Definitions. NIST Special Publication 1500-1. U.S. Letöltés dátuma:

2022.. 11. 21., forrás:<http://dx.doi.org/10.6028/NIST.SP.1500-1>

20. Nogel, M. (2020). A szakértői bizonyítás aktuális kérdései. Budapest: HVG-ORAC.
21. Pancza, J. (2019. 07. 08.). Szövegbányászat. SPSS Nyári Iskola. Veszprém.
22. Sebők, M., Ring, O., & Máté Ákos. (2021). Szövegbányászat és mesterséges intelligencia R-ben. Budapest: Typotex Kiadó.
23. Tikk, D., Farkas, R., Kardkovács, Z., Kovács, L., Répási, T., Szarvas, G., . . . Vázsonyi, M. (2007).
Szövegbányászat. Budapest: Typotex Kiadó.
24. XC. tv. (2005). Az elektronikus információszabadságról.
25. Zódi, Z. (2017). A bírósági ítéletek adatbázisa mint Big Data. Statisztikai Szemle, 95/5.