

Vági Renátó¹

Szemantikai keresés és predictive coding a jogi munkában

*Mit hoz a jövő a jogi adatbázisok
és a jogi keresők szempontjából?*



Absztrakt: Más jogi munkavégzést segítő eszközökhöz hasonlóan a jogi adatbázisok és a jogi keresők is hatalmas fejlődésen mentek át az elmúlt évtizedekben. Azonban míg a jogi informatika általános fejlődéséről az utóbbi időben számos tanulmányt lehet olvasni, addig viszonylag kevesen foglalkoznak az adatbázisok és a keresők fejlődésével. Ebben a tanulmányban ezért az egyes alapfogalmak tisztázásán túl mindenekelőtt azt kívánom bemutatni, hogy hogyan hatott a jogi adatbázisok, valamint keresőmotorok fejlődésére a jogi szövegek formájának és mennyiségének változása. Ezt követően azzal foglalkozom, hogy milyen hiányosságai vannak a jelenleg elterjedt jogi adatbázisoknak, illetve a mesterséges intelligencia, a gépi tanulás, valamint az egyes szövegbányászati eszközök fejlődése hogyan képes segíteni ezek fejlődését. Végül pedig bemutatom, hogy mindennek milyen hatása lehet az olyan költséges és munkaintenzív jogi részfeladatokra, mint a jogi forráskutatás és az elektronikus bizonyítékkeresés, avagy az e-Discovery.

Tárgyszavak: jogi adatbázis; jogi keresőmotor; jogi forráskutatás; e-Discovery; természetesnyelv-feldolgozás (NLP); névelem-felismerés (NER); szemantikai keresés; predictive coding; Technology Assisted Review

Bevezetés

A jogászai munkavégzést segítő különböző technológiai eszközök és szoftverek hatalmas fejlődésen mentek át az elmúlt évtizedekben. Nincs ez máshogy a jogi

¹ Vági Renátó jogász, 2020-ban végzett az Eötvös Loránd Tudományegyetem Állam- és Jogtudományi Karán. Jelenleg az egyetem Jog- és Társadalomelméleti Tanszékének doktorandusz hallgatója. Korábban a Bibó István Szakkollégium hallgatója volt, amelynek jelenleg nevelőtanára. Fő kutatási területe az adatvezérelt technológiák, illetve azok jogra és társadalomra gyakorolt hatásai. Kutatási tevékenysége mellett a MONTANA Tudásmenedzsment Kft. szoftverfejlesztési üzletágának jogi informatikai szakértője.

adatbázisok és a jogi keresők esetében sem. Még mielőtt azonban bemutatnám, hogy milyen fejlődéstörténeten mentek keresztül a jogászai adatbázisok és keresők, illetve milyen fejlődés előtt állnak a közeljövőben, szeretnék tisztázni néhány fogalmat a tanulmány érthetősége érdekében.

Az *adatbázisok* legalapvetőbb meghatározásuk szerint gyűjtemények, amelyek valamilyen adatokat vagy információkat tárolnak valamilyen létrehozott struktúra vagy rendszerszerűség alapján, és amelyek hozzáférhetőek vagy elektronikusan, vagy más módon.² A különböző adatbázisok között pedig különbséget tehetünk egyrészt a tárolt adat tartalma, valamint a tárolás módja szerint.

A *jogi adatbázisok* ez alapján az olyan adatgyűjtemények, amelyek valamilyen szempontból jogilag releváns adatokat tárolnak, vagyis a jogi tudást szeretnék „reprezentálni”, és így elsődlegesen szöveges dokumentumokat tárolnak.³ Ahogy az a későbbiekben látható lesz, ezek az adatbázisok a jogi informatika fejlődése során sokféle formát öltöttek, én pedig alapvetően két típusát különböztetem meg korokon átívelően. Az első az általam a külsőnek, vagyis nyilvánosan elérhetőnek nevezett adatbázisok, amelyek alapvetően a különböző jogforrásokat tartalmazzák. Ezek többnyire általános célú, horizontális szöveges adatbázisok, amelyek egy jogrendszer összes jogi dokumentumát tartalmazzák a teljesség igényére törekedve, és alapvetően sem a felhasználó perbeli szerepére, sem a szakmai területére, sem pedig az eljárás adott fázisára nincs tekintettel.⁴ A másik típus a belső, csupán az adott jogi vállalat (ügyvédi iroda, jogalkotó, jogalkalmazási szerv stb.) számára elérhető adatbázis, amely alapvetően az ügyfelek dokumentumait, iratait, adatait tartalmazza.

Lehet azonban akár milyen jól rendszerezett egy adott adatbázis, egy bizonyos mennyiségű információmennyiség után az adatok valódi hozzáférése és használhatósága csökkenhet. Az egyes adatbázisokhoz tartozó *keresőszolgáltatások* így abban segítenek, hogy az adott információt az adott adatbázis információtengerében minél gyorsabban megtalálja a felhasználó.⁵ Mondhatni, hogy a kereső tulajdonképpen az a felhasználói felület, ahol az ember és a gép interakcióba lép egymással, a felhasználó képes kérdéseket megfogalmazni az algoritmus felé, hogy számára releváns adatokat kapjon.⁶

A *jogi kereső* így funkcionális értelemben nem más, mint az a felület, amely segítséget nyújt a felhasználónak a jogi adatbázis tartalmának feltérképezésében. Technológiai értelemben az adatbázis és a keresőmotor különállnak egymástól, és mindkettő külön elemzés tárgya lehet. Én a továbbiakban egymással összefüggésben értelmezem a kettő fejlődését a jogi adatbázisok és a jogi keresők esetében,

² Zódi Zsolt: *Jogi adatbázisok és jogi forráskutatás*. Budapest, Gondolat Kiadó, 2012. 116–117.

³ Zódi (2012): i. m. 109.

⁴ Zódi (2012): i. m. 107.

⁵ Difference Between Search Engine and Database. Any Difference Between.

⁶ Zódi (2012): i. m. 134.

mert ahogy az a későbbiekben látható lesz, a kettő fejlődése egymás relációjában értelmezhető.

A tanulmány felépítése a következő. *A jogi adatbázisok fejlődése* című részben bemutatom, hogy milyen történeti fejlődésen mentek át eddig a jogi adatbázisok, és milyen technológiai szinten vannak most. Ehhez elsősorban angolszász példákat és angolszász irodalmakat használok fel, aminek alapvetően két oka van. Egyrészt azért, mert az angolszász, azon belül is főleg az egyesült államokbeli jogi informatikai megoldások fejlődési története jobban dokumentált, másrészt pedig azért, mert, ahogy a technológiai fejlődés esetében általánosságban is tapasztalható, a centrumországokban megvalósított technológiai innováció hosszabb vagy rövidebb idő után Magyarországra is begyűrűzik.

Ezt követően *A jogi forráskutatás fejlődésének lehetőségei: névelem-felismerés és a szemantikai keresések* és *Az e-Discovery fejlődésének lehetőségei, avagy a predictive coding* című részekben bemutatom, hogy a jogi adatbázisok előtt milyen fejlesztési lehetőségek állnak, külön kitérve a jogi forráskutatásra és az e-Discoveryre. Végül pedig a *Konklúzió* összefoglalja a tanulmány tartalmát.

A jogi adatbázisok fejlődése

A jogi informatika napjainkig lezajló fejlődését általánosságban három korszakra szokás osztani mind a hazai, mind a nemzetközi szakirodalomban. A jogi informatika az 1970-es években megkezdődött irodaautomatizálási megoldásokkal vette kezdetét. Ezt az első korszakot alapvetően a különböző szövegszerkesztők és táblázatkezelők megjelenése jellemezte, amivel lehetővé vált nagy mennyiségű szövegállományok és adathalmazok egyszerűbb és gyorsabb előállítás, valamint szerkesztése. Ekkor már a nagy mennyiségű szövegállományok miatt megjelennek az első elektronikus jogi adatbázisok, valamint az azt támogató kereső megoldások is. Emellett bizonyos munkafolyamatokat is átláthatóbbá lehetett tenni, valamint automatizálni.⁷ Az ezt követő korszak az elektronikus vagy digitális átállás korszaka. Ebben az időszakban egyre inkább elterjedtek a személyi számítógépek, amelyek teljesítménye és adattárolási képessége egyre jelentősebben növekedett. Emellett az internet megjelenésével és elterjedésével az online kommunikáció, valamint az elektronikus kapcsolattartás is könnyebbé válik, ezáltal egyre inkább eltolódik a jogász munkavégzés a csak elektronikusan tárolt adatok felé, mivel a papíralapú adattárolás és adattovábbítás egyre inkább feleslegessé válik. Ez az az időszak, amikor az első e-Discovery-megoldások is megjelennek.⁸ Jelenleg pedig, nagyjából a 2010-es évek közepétől a harmadik korszakra történő átállás időszakát éljük,

⁷ Zódi Zsolt: A járvány, a jogi szféra és a technológia. *In Medias Res*, 9. (2020), 2. 340.

⁸ Julian Webb: Legal Technology: The Great Disruption. *Melbourne Legal Studies Research Paper Series*, No. 897. 2020. július 31.

amikor egyre több, más iparágakban már elterjedtebb mesterségesintelligencia-alapú megoldás jelenik meg a jogászai munka egyes területein.⁹

A jogi adatbázisok és keresők technológiai fejlődése nagyjából párhuzamosan haladt a jogi informatika általános fejlődésével, ám míg a jogi informatikai fejlesztéseket alapvetően a számítógépek számítási kapacitásának fejlődése, valamint az adattárolási költségek csökkentése határozta meg,¹⁰ addig a jogi adatbázisok fejlődési szükségességének alapvetően a jogi szöveg formája és mennyisége szabott irányt. A jogászai munkavégzés ugyanis alapvetően szöveggéközpontú, úgyhogy az adatbázisokra is azért van szüksége funkcionálisan a jogászoknak, hogy egyrészt a különböző jogforrások szövegei között tudjanak keresni, és a számukra relevánsat megtalálhassák,¹¹ másrészt hogy a munkavégzésük alatt előálló különböző szövegeket (például szerződés, beadvány) tárolni tudják, majd utána egyszerűen képesek legyenek előhívni azokat. Ezért én a továbbiakban, ahogy említettem, különbséget fogok tenni külső, nyilvánosan elérhető adatbázisok és belső, vállalati adatbázisok között. Az előbbi a mindenki számára nyilvános jogszabályok, jogesetek elérését segíti, az utóbbi pedig a jogász munkavégzése során előálló nem publikus adatok tárolására szolgál. Az adatbázisok fejlődésének bemutatása során alapvetően a nyilvános adatbázisok fejlődését fogom bemutatni, aminek két oka van. Egyrészt az, hogy ez szintén jobban dokumentált a szakirodalomban, másrészt a belső adatbázisok kevésbé standardizáltak, hisz a jogász vagy az adott vállalat munkavégzésének sajátosságaira vannak szabva, ezért kevésbé állapíthatók meg általános törvényszerűségek. Ennek ellenére igyekszem majd kitérni a belső adatbázisok sajátosságaira is.

A jogi adatbázisok és a jogi keresők fejlődési történetében szintén három fő korszakot lehet megkülönböztetni, amelyek között két átmeneti korszakról is beszélhetünk. Ezek közül jelenleg a harmadik korszak elején járunk, és a második átmeneti korszakban élünk. Ahogy fentebb is említettem, a jogi adatbázisok fejlődését alapvetően a jogi szövegek mennyisége és formája határozta meg, amit alapvetően az információs krízis fogalmával lehet megragadni.¹² A jogi szövegek információs krízisének a lényege az, hogy bizonyos időszakokban hirtelen megnövekszik a jogi szövegek mennyisége, ami egy idő után nehezen átláthatóvá teszi az addigi formában a szövegállományokat, ezért szükségszerűvé válik egy új típusú formára történő átállás. Ez az új típusú forma azonban azon túl, hogy valóban átláthatóbbá és könnyebben kereshetővé teszi a releváns tartalmakat, egyúttal nem szándékolt hatásként a jogi szövegek előállításának gyorsaságát és így mennyiségét is tovább

⁹ Zódi (2020): i. m. 341.

¹⁰ Daniel M. Katz: Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry. *Emory Law Journal*, 62. (2013), 4. 909–966.

¹¹ Jon Bing: Let There be LITE: a Brief History of Legal Information Retrieval. *European Journal of Law and Technology*, 1. (2010), 1.

¹² Bing (2010): i. m. 5.

növeli, ami így újabb és újabb hatékonyabb formák szükségességét hívja életre. Emellett nyilván nem elhanyagolhatók a jogot körülvevő technológiai, valamint társadalmi és gazdasági változások sem, amelyek szintén az átalakulás irányába mutatnak. A továbbiakban bemutatom az általam azonosított három különböző korszakot, és a közöttük lévő átmeneti korszakokat, valamint kitérek arra, hogy most melyik fázisban tartunk.

Az első korszak nyilvánvaló módon a jogi szövegek hagyományos papíralapú korszaka. Nagyjából a 20. század közepéig ugyanis a jogi szövegek kizárólag papíralapon voltak elérhetők: a jogszabályok és más jogforrások különböző közlönyökben és gyűjteményes művekben jelentek meg, a bírósági határozatok pedig erre specializálódott folyóiratokban és speciális jogeset-gyűjteményekben, a másodlagos jogirodalom pedig a különböző kommentárokból és tudományos folyóiratokban.¹³ A jogi adatbázisok pedig ebben az időszakban tulajdonképpen a könyvtárak volt. Egyrészt a nyilvánosan látogatható könyvtárak, másrészt az egyes személyek, illetve vállalatok személyes könyvtárai, amennyiben megengedhették maguknak, hogy könyveket halmozzanak fel maguknak, valamint volt lehetőségük folyamatosan frissíteni az egyes szövegeket, elkerülve azok elavulását a hatályon kívül helyezés következtében. A belső adatbázisok ebben a korszakban szintén a vállalatok és az egyének belső irattárából és azok rendszeréből álltak. Ezeknek a belső iratoknak a folyamatos rendszerezése, valamint a jogszabályok és az esetjog folyamatos frissen tartása vezetett a jogi könyvtárosi munka felértékelődéséhez.

Már ebben a korszakban megmutatkozott azonban az igény a túlburjánzó jogi dokumentumok rendszerezésére az egyes dokumentumok könnyebb megtalálhatóságának érdekében, különösképpen az esetjog tekintetében a precedensjogi országokban. Ugyanis már a középkortól kezdve készítettek különböző jogeset-gyűjteményeket, amelyek egy-egy terület vagy téma fontosabb bírósági döntéseit tartalmazták különböző szempontok szerint. A jogesetek közlésének forradalmasítását az egyesült államokbeli West jogi kiadóvállalat végezte el a 19. század végén. A többféle reform között például bevezette a Key Number Systemet, amely alá minden egyes ítéletet bekezelte, és így ez egyfajta jogi tezauszává vált az amerikai jogrendszernek. Emellett a West kiadó másfajta indexeket is alkalmazott. A metaadatok új típusa volt a Frank Shepard-féle hivatkozásrendszer is, amelynek segítségével el lehetett különíteni a „fontos” jogeseteket a „kevésbé fontosaktól”.¹⁴ Ezeknek a rendszereknek természetesen az volt a hátránya, hogy csak egy bizonyos szintig szűkítették le a szövegeket (a keresés csak egy meghatározott mélységig volt lehetséges), valamint a különböző helyeken különböző logikával

¹³ Ellie Margolis – Kristen E. Murray: Say Goodbye to the Books: Information Literacy as the New Legal Research Paradigm. *Temple University Legal Studies Research Paper Series, No. 2012-34.* 2012.

¹⁴ Zódi (2012): i. m. 54–60.

készült mutatókhoz esetenként nehéz volt adaptálódni, és ezek gyakran nem tükrözték a tényleges használó logikáját.¹⁵

A jogi adatbázisok fejlődésének második korszaka a 20. század utolsó évtizedében jött el, amikor a személyi számítógépek elterjedésének és az olyan adattároló eszközök, mint a CD feltalálásának és elterjedésének következtében a jogi adatok és jogi szövegek formája egyre inkább eltolódott az elektronikusan tárolt adatok felé.¹⁶ Ez azt a következményt hozta magával, hogy a jogászai munkavégzéshez a szövegek megismerése érdekében nem volt szükség többet könyvtárba járni, vagy saját könyvtárat fenntartani, és azt folyamatosan frissen tartani, hanem mindenkinek lehetősége volt arra, hogy a saját számítógépén vagy CD-n jogi adatbázist vásároljon. Ezt követően pedig a vállalkozás, amelytől a szolgáltatást vásárolta, folyamatosan frissen tartotta számára az adatbázist. A jogi könyvtárosok központi szerepét átvették a szerkesztők. A belső adatbázisok helyzete hasonlóképpen alakult: ezek elkezdtek egyre több belső iratot nem papíralapon, hanem elektronikusan egy saját szerveren tárolni. A keresési lehetőségek azonban még nagyon sokáig a papíralapú keresést utánozták. Ezekben az adatbázisokban a nyomtatotthoz hasonló indexek alapján lehetett keresni.¹⁷ Fontos előrelépés volt, amikor lehetővé vált a szerkesztők által a dokumentumokhoz rendelt kulcsszavak helyett a dokumentum teljes szövegén való szabadszavas keresés, illetve az egyes szavak és kifejezések logikai kombinálására az úgynevezett Boole-féle (Boolean) keresés is.¹⁸ Ez utóbbi azt a lehetőséget foglalja magában, amikor nemcsak egy kulcsszót lehet megadni, hanem akár többet is, és az azok közötti keresési logikai viszonyt be lehet állítani ÉS, VAGY, illetve NEM kategóriákra is.¹⁹ Emellett a legtöbb rendszer lehetővé tette a keresett kifejezésben lévő szó vagy szavak „szótövesítését” a keresés során, ami azt segítette, hogy a felhasználó úgy is megtalálja a keresett dokumentumot, ha abban nem pontosan úgy szerepel a szó, ahogy ő azt a keresés során megadta. A fejlettebb rendszerekben a szavak egymástól való távolságát és a dokumentumban való előfordulási gyakoriságát is be lehetett állítani.²⁰

Az előző század ötvenes-hatvanas éveiben egyre több kutatás foglalkozott a jogi szövegek tárolásának és keresésének hatékonyabbá tételével, ezeket nevezték átfogóan jogi *text retrieval* és *information retrieval* kutatásoknak és eszközöknek.²¹ Végül ezek a kutatások és fejlesztések voltak azok, amelyek eredményeképp létrejöttek a felhasználóbarát, laikus jogászok által is használható jogi adatbázisok.

¹⁵ Margolis–Murray (2012): i. m.

¹⁶ Webb (2020): i. m.

¹⁷ Margolis–Murray (2012): i. m.

¹⁸ Zódi (2012): i. m. 79; Computer-Assisted Legal Research. *Encyclopedia.com*.

¹⁹ Vangie Beal: Boolean Search. *Webopedia*, 2021. május 24.

²⁰ Zódi (2012): i. m. 93.

²¹ Bing (2010): i. m. 2–4.

Az elektronikus jogi adatbázisok széles körű használatát követő átmeneti korszak, amelyben ma is élünk, az internet megjelenésével következett be. Megjelentek, és egyre szélesebb körben elterjedtek az online jogi adatbázisok. Az elmúlt 10-15 évben nemzetközi szinten és itthon is lezajlott az a folyamat a jogi információkkal, amely általánosan végbement az összes információtípussal, és amelynek lényege az, hogy egyre több információ elsődleges megjelenési helye az online tér, illetve egyre több régi dokumentumot is digitalizálnak online.²² A jogi adatbázisok esetében már nem az a kérdés, hogy a jogi szöveg, illetve a keresett jogi információ hol, milyen módon található meg, hiszen lényegében minden fenn van az interneten. A kérdés éppen az, hogy a keresett információt hogyan tudjuk a leggyorsabban és a leghatékonyabban megtalálni. A belső adatbázisok fejlődése ezt a szintet már nem követte, elsősorban az ügyfelek adatbázisban szereplő személyes és érzékeny adatainak biztonsági szempontjai miatt. Az adatbázisokat kiszolgáló keresőmotorok viszont ennek ellenére is további fejlődésen mentek keresztül, ahogy azt a következő pontokban is látni fogjuk.

Az online jogi adatbázisok azonban nagyon sok esetben még nem tértek át az új típusú forma által megkívánt és lehetővé is tett keresési megoldásokra. Az online adatbázisok ugyanis sokáig többnyire csak az elérhető dokumentumok folyamatos bővítésével és a dokumentumok közötti súlyozási rendszer bevezetésével adtak hozzá pluszt a korábbi adatbázisok keresési lehetőségeihez. Utóbbi azt jelenti, hogy a találati listában megjelenő dokumentumokat különböző szempontok szerint, például a keresett kifejezés előfordulási gyakorisága vagy előfordulási helye, az adott dokumentumra történő hivatkozás gyakorisága szerint rangsorolta a rendszer előrébb vagy hátrébb.²³ Ezért a harmadik korszak betetőzéséről akkor beszélhetünk, ha a jogi adatbázisok után az azokban történő keresés is az online térre lesz szabva. Az adat újabb típusú elérési módja ugyanis új típusú keresési és adatfeldolgozási módokat, valamint új típusú összefüggések felderítését teszi lehetővé az egyes jogi dokumentumok között. Ezért azt is mondhatjuk, hogy a szerkesztők munkáját egyre inkább kiegészíti a fejlesztő, illetve a *data scientist* munkája.

A következő pontokban azt mutatom meg, hogy a közeljövőben milyen új típusú fejlesztések várhatók mind a külső, nyilvános adatbázisok, amelyeket elsősorban a jogi forráskutatáshoz használnak fel, mind pedig a belső, saját vállalati adatbázisok esetében, amelyet viszont elsősorban az e-Discoveryhez használnak fel. Azért tartom fontosnak a jogi forráskutatás és az e-Discovery középpontba állításával bemutatni a két típusú adatbázis fejlődési lehetőségeit, mert konkrét tapasztalati példákon talán érthetőbbé válik ezeknek az új típusú keresési módoknak a fontossága. Másrészt pedig azért, mert mind a jogi forráskutatás, mind az e-Discovery a jogi munkavégzésnek az a része, amelyre különböző nemzetközi kutatások alapján a legtöbb időt és a legtöbb emberi erőforrást áldozzák a jogi szervezetek, így ezek hatékonyabbá

²² Margolis–Murray (2012): i. m.

²³ Zódi (2012): i. m. 137.

tételével és bizonyos módon automatizálásával van lehetőség a jogász munkát is hatékonyabbá tenni.

A jogi forráskutatás fejlődésének lehetőségei: névelem-felismerés és a szemantikai keresések

Érdekes módon a „jogi forráskutatás” angol szakkifejezést, a *legal research*-öt az angol-szász jogi kultúrában alapvetően kettős értelemben használják. Egyrészt lefedi az itthon jogi forráskutatásnak nevezett folyamatot, tehát az egyedi jogi ügyben szükséges jogi felkészülési folyamatot, amelyet alapvetően a praktizáló jogászok végeznek. Másrészt viszont olykor utalnak vele az akadémiai jogi kutatásokra is. Ezek, az előzővel ellentétben inkább bizonyos empirikus és természettudományos módszereket használnak a jog területén, például a jogszabályok és jogalkalmazási szervek által hozott határozatok statisztikai vizsgálatát, tartalomelemzését.²⁴ Ez utóbbit itthon inkább a jogszociológia területére szoktuk sorolni. A következőkben bemutatott megoldások alapvetően a jogi forráskutatás támogatására érvényesek, viszont nyilvánvalóan bizonyos megoldások az empirikus kutatások területén is használhatók.

A jogi forráskutatás jelen tanulmányban használt fogalma tehát a jogász munkafolyamatnak az a jól körülhatárolható része, amelynek során a jogászok összegyűjtik az aktuális ügyük szempontjából releváns és hatályos jogi rendelkezéseket. Ez az adott ügy bármelyik szakaszában előfordulhat, tehát akkor is, ha csak egyszerű tanácsadással szolgálnak az ügyfelük részére, vagy csak egy szerződést fogalmaznak meg, de nyilvánvalóan akkor van ennek a legnagyobb szerepe, amikor bíróság előtti pereskedésre készülnek, és azokat a jogi érveket keresik, hogy a bíróság miért az ő ügyfelük javára döntsön.²⁵ Ez elsősorban itthon a magyar jogi kultúra jellemzői miatt alapvetően jogszabályok kereséséből áll, de nyilvánvalóan nem elhanyagolhatók a bírósági határozatokra vagy más jogalkalmazó szervek ítéleteire, illetve a jogirodalomra történő keresések sem.

A jogi forráskutatás ma már alapvetően a nyilvános jogi adatbázisok igénybevételel történik, azonban ahogy az korábban szóba került, az ezekben való kulcsszavas vagy összetett kulcsszavas (Boolean) típusú keresések sok esetben nem hatékonyak, ami ezt a típusú munkavégzést időigényessé teszi. A jogi adatbázisokkal foglalkozó

²⁴ Bart Custers: *Methods of Data Research for Law*. In Vanessa Mak – Eric Tjong Tjin Tai – Anna Berlee (szerk.): *Research Handbook in Data Science and Law*. Research Handbooks in Information Law. Cheltenham, Edward Elgar, 2018. 364.

²⁵ Jake Heller – Pablo Arredondo: *AI in Legal Research: How AI Is Providing Everyone Access to Information and Leveling the Playing Field for Firms of All Sizes*. In Noah Waisberg – Alexander Hudek (szerk.): *AI For Lawyers: How Artificial Intelligence is Adding Value, Amplifying Expertise, and Transforming Careers*. Hoboken, New Jersey, John Wiley & Sons, 2021.

szakirodalom ezt a problémát az under-inclusivity (túlszűrés), az over-inclusivity (alulszűrés), valamint a homogenitás fogalmak hármásával írja le. Túlszűrésről akkor beszélünk, ha a keresés elküldését követően a találati lista nem tartalmaz minden olyan dokumentumot, amely számunkra releváns lehet. Ez egyrészt előfordulhat azért, mert túl specifikus keresési feltételt adtunk meg, másrészt pedig azért, mert a különböző jogi szövegekben nem feltétlenül következetesen használják ugyanazt a jogi fogalmat. Ezzel szemben alulszűréskor a találati lista sok olyan dokumentumot tartalmaz, amely számunkra nem releváns, csak zajt kelt, és időpazarlást okoz. A pontatlan találatok miatt ezért a jogi forráskutatás nagyon időigényes tevékenységgé válhat. Ezzel van összefüggésben a homogenitás problémája is: ha a találati listában az egyes jogi dokumentumok esetében csak azok megnevezése, ügyszáma vagy csak az érintett résztvevők megnevezése szerepel, az a jogi forráskutatást végző személy számára nem ad sok információt a dokumentum tartalmáról, ezért azt minden esetben meg kell nyitnia, és át kell futnia a szövegét.²⁶

Ezekre a problémákra nyújtanak megoldást a különböző természetes nyelvfeldolgozást, vagyis *natural language processinget* (a továbbiakban: NLP) lehetővé tevő megoldások. Az NLP a mesterséges intelligencia gépi tanulás melletti másik olyan részterülete, amely releváns a jogi informatika szempontjából. Röviden megfogalmazva, természetes nyelvfeldolgozásnak azokat a szoftveres megoldásokat nevezzük, amelyek képesek a strukturálatlan szövegeket, azok statisztikai törvényszerűségeit felismerve különbözőképpen feldolgozni. Az ilyen típusú megoldásokat jelenleg a nyelvfelismerésben, a gépi fordításban, az automatikus összefoglaló készítéséhez, valamint az automatikus érzelemfelismeréshez szokták hatékonyan használni.²⁷ Az utóbbi időben széles körben elterjedt megoldások azon az egyszerű felismerésen alapszanak, hogy a jogi nyelv is minden specialitása ellenére a természetes nyelvek egy csoportja. Vagyis természetes nyelvfeldolgozó eszközökkel a strukturálatlan jogi szövegeket is át lehet úgy alakítani, hogy az a gépek számára értelmezhető és különböző szempontok szerint elemezhető legyen.²⁸

Az NLP egyik meghatározó felhasználási területe a jogi szövegek esetében az információkinyerés (*information extraction* – IE). Az IE meghatározott információelemek megtalálása és kinyerése folyó szövegekből. Ezeknek a megoldásoknak a lényege az, hogy ha ismerjük az adott információtípus szabályszerűségeit a szövegen belül, akkor ezeket a szabályszerűségeket megfogalmazva automatikusan megjelölhetővé és kiemelhetővé, külön tárolhatóvá válnak a szövegből ezek az elemek,

²⁶ Ed Walters – Jeff Asjes: Fastcase, and the Visual Understanding of Judicial Precedents. In Daniel Martin Katz – Ron Dolin – Michael J. Bommarito (szerk.): *Legal Informatics*. Cambridge, Cambridge University Press, 2021. 358–359.

²⁷ Daniel Martin Katz: AI + Law. In Daniel Martin Katz – Ron Dolin – Michael J. Bommarito (szerk.): *Legal Informatics*. Cambridge, Cambridge University Press, 2021. 89.

²⁸ John J. Nay: Natural Language Processing for Legal Texts. In Daniel Martin Katz – Ron Dolin – Michael J. Bommarito (szerk.): *Legal Informatics*. Cambridge, Cambridge University Press, 2021. 99–100.

majd bármilyen későbbi felhasználásra elérhetővé tehető.²⁹ Az információkinyerés jogi szövegek esetében legrelevánsabb területe pedig a névelem-felismerés, vagyis a *named entity recognition* (NER). Ebben az esetben a felismerendő és kinyerendő információtartalom valamilyen elnevezés,³⁰ jogi szövegek esetében legtöbbször jogszabály-hivatkozás, bírósági ítéletre történő hivatkozás, dátum, a peres felek, illetve képviselőik és a bírák neve, de akár körülhatárolható jogi fogalmak is. A jogi adatbázisok esetében nyilvánvalóan ezek a megoldások azért relevánsak, mert ezek segítségével automatikusan, szerkesztői segítség vagy közbeavatkozás nélkül kinyerhetővé válnak jogi információ entitások, amelyek utána kereshetővé tehetőek. Így az adatbázist használó jogász felhasználónak a kulcsszavas keresésnél több lehetősége van a szűrésre, így könnyebben találhatja meg a számára releváns jogi dokumentumot. Emellett pedig a különböző hivatkozások kinyerésével a jogi dokumentumok közötti kapcsolat is sokkal láthatóbbá válik, így ha a felhasználó talál egy számára releváns dokumentumot, akkor utána nem kell egy új keresést lefuttatnia az ahhoz hasonló dokumentumok megtalálása érdekében, hanem egyszerűen átválthat az eredeti dokumentumhoz kapcsolódó más dokumentumokhoz.

A természetes nyelvfeldolgozó megoldások másik felhasználási területe a jogi szövegek esetében a szemantikai feldolgozás lehetővé tétele. A szemantikai megoldások alapvetően abban különböznek az információkinyeréstől és a névelem-felismeréstől, hogy nem a szövegben benne lévő konkrét információelemet nyernek ki és tesznek láthatóvá és/vagy kereshetővé, hanem a szöveg teljes egészének elemzésére adnak lehetőséget jelentéstani szempontból. Ezzel nem a szövegek közötti összefüggéseket teszik láthatóvá, hanem az egymáshoz hasonló dokumentumok megtalálását könnyebbé. Az ilyen típusú megoldások a jogi területen elsősorban a különböző dokumentumok automatikus kategorizálására hasznosíthatók.³¹ Ez ugyanis lehetővé teszi, hogy a különböző jogi dokumentumok ne csak egyféle szempont szerint (például az aktuális jogi dogmatika szabályszerűségei szerint) legyenek csoportosítva, hanem más összefüggések szerint is. Emellett nemcsak egy-egy jogforráson belül teszi ezt lehetővé, hanem a különböző jogforrások között is átjárhatóvá teszi az utat és a keresést. Jelenleg ugyanis a legtöbb jogi adatbázisban mind az elsődleges jogforrásokat, mint például a jogszabályokat, bírósági határozatokat stb., mind pedig az ezeket magyarázó tudományos jogirodalmat külön kezelik az adatbázis jobb átláthatósága miatt. A különböző szemantikai megoldások viszont lehetővé tudják

²⁹ Neda Sakhaee – Mark C. Wilson: Information Extraction Framework to Build Legislation Network. *Artificial Intelligence and Law*, 29. (2021), 1. 37.

³⁰ Sakhaee–Wilson (2021): i. m. 38.

³¹ Irina Matveeva: Predictive Coding in E-Discovery. In Daniel Martin Katz – Ron Dolin – Michael J. Bommarito (szerk.): *Legal Informatics*. Cambridge, Cambridge University Press, 2021. 323.

tenni, hogy a jogirodalom integrált módon legyen kereshető a többi jogforrással közösen.³²

A szemantikai feldolgozások másik termékeny területe a jogi forráskutatás területén az összefoglalók automatikus elkészítése.³³ Ahogy ugyanis említettem, a jogi keresők egyik nagy hiányossága az úgynevezett homogenitás, tehát hogy a találati listában nem sok információ jelenik meg az egyes dokumentumokról, így nehezen eldönthető annak relevanciája az adott felhasználó számára. A bírósági határozatok esetében ezen mindenféleképpen segíteni tud, ha egy, az adott ügy összes lényegi elemét összegyűjtő, érthető összefoglaló olvasható el róluk gyorsan, ami alapján rövid idő alatt eldönthetővé válik, hogy érdemes-e számunkra az az adott dokumentum, vagy nem.

Természetesen még nagyon hosszan lehetne sorolni, hogy milyen különböző NER-alapú és szemantikai megoldások válhatnak majd elérhetővé a jogi forráskutatás számára a jövőben, azonban jelen tanulmány célja, hogy a közelmúlt és a közeljövő megoldásait mutassa be. Ezért a továbbiakban áttérek az e-Discoveryre mint a jogászai munkavégzés másik nagy időigényű munkafolyamatára, azon belül is a *predictive coding*-ra, ami ennek a típusú munkavégzésnek a hatékonyságát növeli.

Az e-Discovery fejlődésének lehetőségei, avagy a predictive coding

A jogi forráskutatás mellett a jogi munkafolyamatok másik legidőigényesebb fajtája az e-Discovery, amelyet magyarul elektronikus bizonyítékkeresésnek nevezhetünk. Az e-Discovery főleg az angolszász jogi kultúrában használt kifejezés, amely a jogászok munkájának azt a részét írja le, amely az akármilyen üzyszakban lévő peres ügyének érdekében az elektronikus adatok keresését, rendszerezését, elemzését, valamint bemutatását foglalja magában annak érdekében, hogy az ügyfél számára a jogi képviselő előnyös bizonyítékot tárhasson a bíróság elé.³⁴ Mivel a peres eljárások bizonyítékai jelentős részben szövegállományok, és az utóbbi időben egyre nagyobb részben elektronikusan tárolt szöveges dokumentumok, ezért mind maga az e-Discovery, mind az e-Discoveryt támogató megoldások fontossága egyre nagyobb. Az e-Discovery térhódításának kezdetét nagyjából a 2000-es évek közepére tehetjük egyrészt azért, mert ekkor kezdett teret nyerni a céges kommunikációban az e-mailezés, amely komoly mértékű elektronikusan tárolt, és egy peres ügy számára

³² Enrico Francesconi – Ginevra Peruginelli: Integrated Access to Legal Literature Through Automated Semantic Classification. *Artificial Intelligence and Law*, 17. (2009), 1. 31–49.

³³ Emmanuel Chieze – Atefeh Farzindar – Guy Lapalme: An Automatic System for Summarization and Information Extraction of Legal Information. In Enrico Francesconi et al. (szerk.): *Semantic Processing of Legal Texts*. Berlin, Springer, 2010. 216–234.

³⁴ Jack G. Conrad: E-Discovery Revisited: The Need for Artificial Intelligence Beyond Information Retrieval. *Artificial Intelligence and Law*, 18. (2010), 4. 322.

potenciálisan releváns szöveget generált.³⁵ Másrészt a 2000-es évek közepén engedélyezte az Egyesült Államok szövetségi polgári peres eljárásait szabályozó kódexe (US Federal Rules of Civil Procedure) az elektronikus dokumentumok bizonyítékként történő felhasználását.³⁶

Az Egyesült Államokban különböző e-Discoveryvel foglalkozó gyakorló jogászok és az e-Discoveryt kutató egyetemek és más kutatói helyek közösen megalkottak egy modellt, amely az e-Discovery egyes lépéseit és azok helyes módját írja le az adat életciklusának középpontba állításával. Ezt a modellt Electronic Discovery Reference Modelnek, vagyis EDRM-nek nevezték el, és az évek folyamán folyamatosan alakították, ráadásul mára már egyszerre több modell is él párhuzamosan, habár általánosságban a lényegük azonos.³⁷ Ahogy az 1. ábrán is látszik, ezek a modellek az *information governance*-től, tehát az információk vállalaton belüli kezelésétől egészen a prezentációig, tehát a bizonyítékok bíróság elé tárásáig követik nyomon, hogy mit és hogyan lehet kezdeni az adatokkal, és ezen belül főleg a szöveges adattal. Emellett ezeket a különböző eljárásokat alapvetően két részre szokták osztani, az ábra bal részére, ami az adatok kezelésének technikai oldalát mutatja, mint például az adatok begyűjtésének és tárolásának a módját, amely inkább az informatikai területét írja le az e-Discoverynek, és az ábra jobb részére, amely inkább a jogi munkát ábrázolja, mint az adatok feldolgozása, átnézése, elemzése és prezentációja.³⁸ Habár bizonyos szempontok szerint az első rész sem elhanyagolható jogász szemmel, ugyanis eléggé meghatározza a későbbi munkavégzés módját, én azonban a továbbiakban mégis csupán a második résszel fogok foglalkozni, ugyanis a jogász munkavégzés szempontjából ez az, ami relevánsabb.

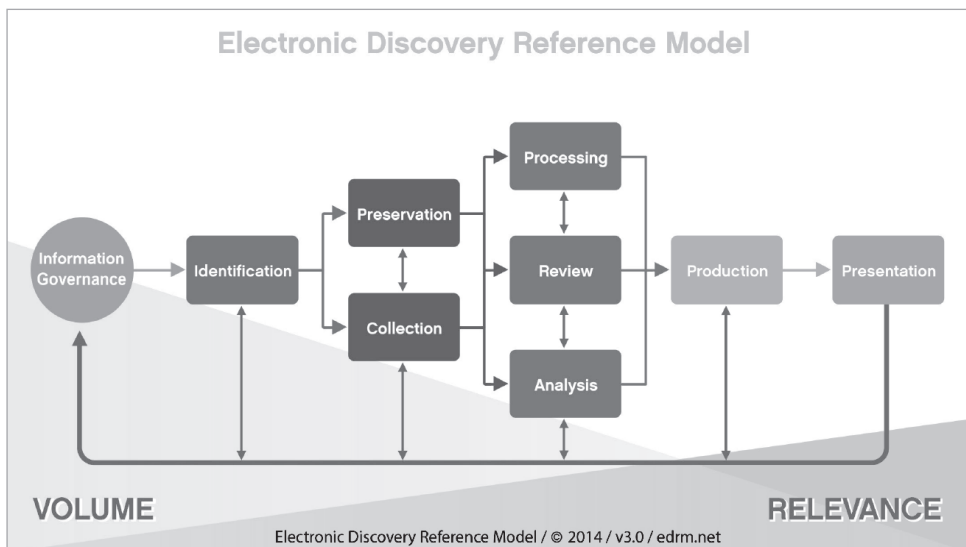
Az e-Discoveryt támogató jogi informatikai eszközök tehát alapvetően az adatok feldolgozását, az adatok elemzését és a megoldások bemutatását tartalmazó munkafolyamatokat tudják hatékonyan támogatni. Az adatok feldolgozása, vagyis a processing tulajdonképpen az adatok kereshetővé tétele és az információk közötti keresés. Így az alapvető probléma itt is nagyon hasonló, mint a jogi forráskutatást segítő nyilvános adatbázisoknál: a kulcsszavas keresésre építő keresőmotorok sok esetben nem adnak megfelelő találatokat, vagy a releváns találatok megtalálásának idejét nem csökkentik kellőképpen. Így az előző részben bemutatott névelemfelismerő alapú és szemantikai megoldások a belső adatbázisok találati listájának hatékonyságát is nagyban tudják segíteni.

³⁵ Dera Nevin: eDiscovery: What It Is and How AI Is Continuing to Transform How It Works for You. In Noah Waisberg – Alexander Hudek (szerk.): *AI For Lawyers: How Artificial Intelligence is Adding Value, Amplifying Expertise, and Transforming Careers*. Hoboken, New Jersey, John Wiley & Sons, 2021.

³⁶ Matveeva (2021): i. m. 315.

³⁷ Jonathan Kerry-Tyerman – A. J. Shankar: The Core Concepts of E-Discovery. In Daniel Martin Katz – Ron Dolin – Michael J. Bommarito (szerk.): *Legal Informatics*. Cambridge, Cambridge University Press, 2021. 291–292.

³⁸ Kerry-Tyerman–Shankar (2021): i. m. 292.



1. ábra: Az elektronikus bizonyítékok kutatását bemutató és leíró modell³⁹

Megjegyzés: Electronic Discovery Reference Model: Elektronikus bizonyítékkeresés referencia modellje; Information Governance: Szervezeti információs stratégia; Identification: Azonosítás; Preservation: Tárolás; Collection: Gyűjtés; Processing: Feldolgozás; Review: Átnézés; Analysis: Elemzés; Production: A kész anyag elkészítése; Presentation: Bemutatás; Volume: Mennyiség; Relevance: Relevancia

Forrás: Frameworks & Standards. *EDRM*.

A másik része az e-Discovery munkafolyamatainak az egyes dokumentumok közötti keresésen túl, amelyet a mesterséges intelligenciával megtámogatott belső adatbázisok és belső jogi keresők tudnak segíteni és bizonyos részben automatizálni, az a releváns bizonyítékok átnézése és elemzése. Valamilyen szinten természetesen ez is kapcsolódik a kereséshez, ugyanis az első lépcsője a releváns adatok megtalálása, ám bizonyos értelemben túlmutat azon. A bizonyítékok gyorsabb átnézését és elemzését lehetővé tevő megoldásokat összefoglaló néven *technology assisted review*-nek (TAR) vagy más elnevezéssel *predictive coding*-nek szokták nevezni. Ezeknek a megoldásoknak a lényege, hogy egy-egy ügytípuson belül a különböző egyéni esetekben nagyrészt ugyanazok vagy legalábbis nagyon hasonló dokumentumok és információk lesznek relevánsak a jogi eljárásban. Így, ha egy jogásznak már volt megfelelő mennyiségű korábbi hasonló esete, mint amilyenrel újonnan találkozok, akkor egy mesterségesintelligencia-alapú algoritmust meg lehet tanítani arra, hogy automatikusan ajánlja fel az olyan típusú és tartalmú dokumentumokat a jelen ügyben, mint amelyek a korábbiakban relevánsak voltak.⁴⁰ Az „előrejelző kódolás” kifejezés így

³⁹ Frameworks & Standards. *EDRM*.

⁴⁰ Nevin (2021): i. m.

tulajdonképpen azt takarja, hogy a gép meg tudja jelölni azokat a dokumentumokat, amelyek nagy eséllyel fontosak lesznek az adott ügyben.

Az ilyen típusú algoritmusok előállításának jelenleg két elterjedt formája van. Az egyik az úgynevezett egyszerű passzív tanulás, amikor szakértők meghatározzák az adott ügýtípusban releváns dokumentumok körét, amelyet tanító adatnak használnak fel a *predictive coding* algoritmus előállításához. Ha pedig az algoritmus elér egy bizonyos mértékű elfogadható pontosságot, akkor elkezdik élesben is használni. A másik fajta előállítási mód a folyamatos tanulás, amikor az adott vállalat jogászai folyamatosan címkézik a munkavégzésük során a releváns dokumentumokat/dokumentumrészleteket és információkat, az algoritmus pedig ezek alapján folyamatosan tanul, és folyamatosan képes fejleszteni magát az emberi címkézők visszajelzései alapján.⁴¹

Az e-Discovery folyamat vége pedig a talált bizonyítékok bemutatása akár az ügyfelek részére, akár a bíróság számára. A különböző mesterségesintelligencia-megoldások ebben is segítséget nyújtanak a jogászoknak, ugyanis egyrészt, ha a bizonyítékok felkutatására és elemzésére kevesebb időt lehetett szánni, akkor több idő marad a prezentáció elkészítésére, ami befogadhatóbbá tudja tenni azt. Másrészt pedig sok e-Discoveryt támogató MI-megoldás képes az egyes dokumentumok és információk közötti kapcsolatok bemutatására, ami sokkal átláthatóbbá tehet hatalmas bizonyítékmennyiségeket.⁴²

Ezek a megoldások természetesen elsősorban az ügyvédek (akár büntető, akár nem büntető ügyszakban) és az ügyészek munkafolyamatait tudják megkönnyíteni, azonban a bíróságok számára sem lényegtelenek ezek. Egyrészt azért, mert bizonyos eljárások esetében a bíróságok is szenvednek a nagy mennyiségű adathalmaztól, amelyeket az eljáró felek bizonyítékként eléjük tárnak. Ezekben az esetekben nekik is segítséget nyújthatnak azok a predictive coding megoldások, amelyek láthatóbbá teszik számukra az ügy szempontjából nagy eséllyel releváns információkat, amivel a saját ügyterhüket és a perek elhúzódásának az esélyét tudják csökkenteni. Emellett a különböző prezentációs, az összefüggéseket láthatóbbá tevő eszközöket is alkalmazhatnak, amelyekkel a döntéseik indokolását tehetik közérthetőbbé, ezáltal pedig a joghoz való hozzáférés és a döntéseik legitimációját növelhetik. Végül pedig a bíróságoknak szerepük lehet abban is, hogy a különböző *predictive coding*, illetve más mesterségesintelligencia-megoldások peres eljárásokban történő alkalmazásához támpontokat és szempontrendszereket adjanak akár a gyakorló jogászoknak, akár szélesebb körben az állampolgároknak, ahogy azt az Egyesült Államokban a *Da Silva Moore v. Publicis Groupe* ügyben tették a bíróságok.⁴³ Ez ugyanis nagymértékben növelheti az új technológiába vetett bizalmat, és azok szélesebb körű elterjedését. Emellett pedig ez az alkalmazás használható lehet a hasonló bírósági

⁴¹ Kerry-Tyerman-Shankar (2021): i. m. 308.

⁴² Nevin (2021): i. m.

⁴³ *Electronic Discovery Update*. (2012), Fall. 2.

ítéletek megtalálására is, ami a korlátozott precedensrendszerre való átállás miatt egyre nagyobb fontosságú lehet a jövőben.

Konklúzió

Összefoglalva tehát a fentebb leírtakat, technológiai szempontból az adatbázisok és a keresők jelenleg átmeneti fázisban vannak az elektronikus keresést lehetővé tevő jogi adatbázisok és az online keresést lehetővé tevő adatbázisok között. Az előálló jogilag releváns dokumentumok ugyanis nagyrészt már online formában elérhetők, a keresési lehetőségek azonban sok esetben még nem adaptálódtak ehhez az új típusú formátumhoz. A jogi munkavégzésnek a jogi forráskutatási, valamint az elektronikus bizonyítékok megkeresését és elemzését célzó munkafolyamatai így még mindig meglehetősen idő- és emberierőforrás-igényes feladatok. Egy létező jogi szervezet ezért jelentős költségeket tud megtakarítani, egy, a piacra újként belépni kívánó ügyvédi iroda pedig versenylőnyre tehet szert, amennyiben különböző mesterségesintelligencia-megoldásokkal képes hatékonyabbá tenni és bizonyos fokon automatizálni az említett a folyamatokat. Segítséget nyújtanak a nyilvánosan elérhető adatbázisok esetében a különböző névelem-felismerő és szemantikai alapú megoldások, a belső adatbázisoknál pedig ezeken túl az e-Discoveryt segítő *predictive coding*. A tanulmányban is bemutatott megoldások általában vagy már elérhetők, ha nem is feltétlenül magyarul, vagy hamarosan nagy eséllyel elérhetővé válnak a jogászok számára.

Felhasznált irodalom

- Beal, Vangie: Boolean Search. *Webopedia*, 2021. május 24. Online: www.webopedia.com/definitions/boolean-search/
- Bing, Jon: Let There be LITE: a Brief History of Legal Information Retrieval. *European Journal of Law and Technology*, 1. (2010), 1. Online: <https://ejlt.org/index.php/ejlt/article/view/15>
- Chieze, Emmanuel – Atefeh Farzindar – Guy Lapalme: An Automatic System for Summarization and Information Extraction of Legal Information. In Enrico Francesconi – Simonetta Montemagni – Wim Peters – Daniela Tiscornia (szerk.): *Semantic Processing of Legal Texts*. Berlin, Springer, 2010. 216–234. Online: https://doi.org/10.1007/978-3-642-12837-0_12
- Computer-Assisted Legal Research. *Encyclopedia.com*. Online: www.encyclopedia.com/law/encyclopedias-almanacs-transcripts-and-maps/computer-assisted-legal-research
- Conrad, Jack G.: E-Discovery Revisited: The Need for Artificial Intelligence Beyond Information Retrieval. *Artificial Intelligence and Law*, 18. (2010), 4. 321–345. Online: <https://doi.org/10.1007/s10506-010-9096-6>
- Custers, Bart: Methods of Data Research for Law. In Vanessa Mak – Eric Tjong Tjin Tai – Anna Berlee (szerk.): *Research Handbook in Data Science and Law*. Research Handbooks

- in Information Law. Cheltenham, Edward Elgar, 2018. 355–377. Online: <https://doi.org/10.4337/9781788111300.00023>
- Difference Between Search Engine and Database. *Any Difference Between*. Online: <https://anydifferencebetween.com/difference-between-search-engine-and-database/>
- Electronic Discovery Update*, (2012), Fall. Online: www.kramerlevin.com/images/content/1/4/v4/1498/2012-0901-Electronic-Discovery-Update-Newsletter.pdf
- Frameworks & Standards. *EDRM*. Online: <https://edrm.net/resources/frameworks-and-standards/>
- Francesconi, Enrico – Ginevra Peruginelli: Integrated Access to Legal Literature Through Automated Semantic Classification. *Artificial Intelligence and Law*, 17. (2009), 1. 31–49. Online: <https://doi.org/10.1007/s10506-008-9072-6>
- Heller, Jake – Pablo Arredondo: AI in Legal Research: How AI Is Providing Everyone Access to Information and Leveling the Playing Field for Firms of All Sizes. In Noah Waisberg – Alexander Hudek (szerk.): *AI For Lawyers: How Artificial Intelligence is Adding Value, Amplifying Expertise, and Transforming Careers*. Hoboken, New Jersey, John Wiley & Sons, 2021.
- Katz, Daniel M.: Quantitative Legal Prediction—or—How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry. *Emory Law Journal*, 62. (2013), 4. 909–966. Online: <https://scholarlycommons.law.emory.edu/elj/vol62/iss4/6>
- Katz, Daniel Martin: AI + Law. In Daniel Martin Katz – Ron Dolin – Michael J. Bommarito (szerk.): *Legal Informatics*. Cambridge, Cambridge University Press, 2021. 87–93. Online: <https://doi.org/10.1017/9781316529683>
- Kerry-Tyerman, Jonathan – A. J. Shankar: The Core Concepts of E-Discovery. In Daniel Martin Katz – Ron Dolin – Michael J. Bommarito (szerk.): *Legal Informatics*. Cambridge, Cambridge University Press, 2021. 291–314. Online: <https://doi.org/10.1017/9781316529683.021>
- Margolis, Ellie – Kristen E. Murray: Say Goodbye to the Books: Information Literacy as the New Legal Research Paradigm. *Temple University Legal Studies Research Paper Series*, No. 2012–34., 2012. Online: <https://doi.org/10.2139/ssrn.2125278>
- Matveeva, Irina: Predictive Coding in E-Discovery. In Daniel Martin Katz – Ron Dolin – Michael J. Bommarito (szerk.): *Legal Informatics*. Cambridge, Cambridge University Press, 2021. 315–334.
- Nay, John J.: Natural Language Processing for Legal Texts. In Daniel Martin Katz – Ron Dolin – Michael J. Bommarito (szerk.): *Legal Informatics*. Cambridge, Cambridge University Press, 2021. 99–113. Online: <https://doi.org/10.1017/9781316529683.011>
- Nevin, Dera: eDiscovery: What It Is and How AI Is Continuing to Transform How It Works for You. In Noah Waisberg – Alexander Hudek (szerk.): *AI For Lawyers: How Artificial Intelligence is Adding Value, Amplifying Expertise, and Transforming Careers*. Hoboken, New Jersey, John Wiley & Sons, 2021.
- Sakhaee, Neda – Mark C. Wilson: Information Extraction Framework to Build Legislation Network. *Artificial Intelligence and Law*, 29. (2021), 1. 35–58. Online: <https://doi.org/10.1007/s10506-020-09263-3>
- Walters, Ed – Jeff Asjes: Fastcase, and the Visual Understanding of Judicial Precedents. In Daniel Martin Katz – Ron Dolin – Michael J. Bommarito (szerk.): *Legal Informatics*. Cambridge, Cambridge University Press, 2021. 357–370. Online: <https://doi.org/10.1017/9781316529683.024>
- Webb, Julian: Legal Technology: The Great Disruption. *Melbourne Legal Studies Research Paper, Series No. 897*. 2020. július 31. Online: <https://doi.org/10.2139/ssrn.3664476>
- Zódi Zsolt: A járvány, a jogi szféra és a technológia. *In Medias Res*, 9. (2020), 2. 339–355.
- Zódi Zsolt: *Jogi adatbázisok és jogi forráskutatás*. Budapest, Gondolat Kiadó, 2012.