

Pécsi Tudományegyetem
Állam- és Jogtudományi Kar
Doktori Iskola

Zódi Zsolt

Jogi adatbázisok és jogi forráskutatás

Gépek a jogban

Doktori értekezés tézisei
Témavezető:
Dr. Balogh Zsolt György PhD
tanszékvezető egyetemi docens

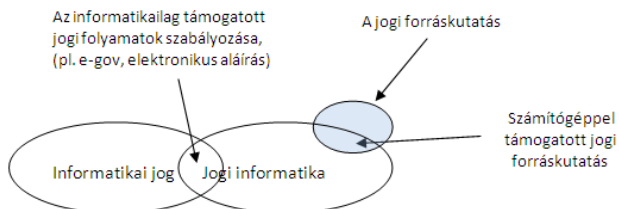
I. A kitűzött kutatási feladat

A számítógépet meglepően régóta használják a jogi munkában. Az ötvenes évek végének Amerikájában a jogi szövegek feldolgozására és tárolására nem is egy projekt indult, amelyek közül néhány már a hetvenes évekre üzleti alapokon működő, széles körben használt rendszerré vált. Ez azt is jelenti, hogy a jogászok sokkal korábban használtak elektronikus adatbázisokat és keresőket, mint bármely más szakma. A magyar jogászokra is igaz, hogy évekkal korábban mozogtak otthonosan a digitális szövegek és keresők világában, mint ahogy például a *Google* megjelent.

A számítógép és a jog találkozási pontján ma már rengeteg témát, problémát, és legalább két tudományterületet találhatunk: az informatikai jogot és a jogi informatikát. Előbbi, a *computer law* már a hatvanas – hetvenes években egészen szilárd tematikával rendelkezett, és az Internet ugyan hozott egy sor új szempontot, mégis azt mondhatjuk, hogy egy patinás múlttal rendelkező jogterületről van szó. Ezt a területet jogászok művelik. Utóbbit - a jogi informatikát - viszont inkább mérnökök, és több részterületből áll, amelyek közül a *jogi adatbázisok* témaköre egyetlen, nem is túl nagy, és talán nem is a legfontosabb. Érintkezésben van az ún. *text retrieval* (szövegkinyerés) és a *database design* (adatbázis-tervezés) témaköreivel, és érintkezik az *artificial intelligence* (mesterséges intelligencia) és az *information science* (információtudomány) tematikával, amely utóbbi pedig a hagyományos *library science* (könyvtártudomány) részének (is) tekinthető.

A *legal research* (jogi forráskutatás) témát Amerikában voltaképp a kiadói szakma teremtette meg, és ma elsősorban

jogi könyvtárosok, és információ-menedzserek gondozák: dominálnak benne az információ-szervezési, könyvtár-tudományi, informatikai, és praktikus-didaktikai szempontok. Bár magam is kiadói ember voltam, és a jogi adatbázisok, a jogi forráskutatás, és a keresők használata a napi munkámban folyamatos fejtörést jelentő kérdések voltak éveken át, ez a dolgozat valamivel ambiciózusabb annál, hogy egyszerűen csak a kereséssel, vagy a jogi forráskutatással, esetleg a jogi adatbázisok történetével foglalkozzon. A fenti tudományterületek mindegyikéről nyert inspirációt, és a specialitása, hogy ezt erős elméleti háttér elé helyezi. Azt hiszem, ezt a látásmódot szokták manapság eléggé elhasznált szóval interdiszciplinárisnak mondani.



Ha elkezdjük ezzel a személetmóddal vizsgálni a jog és a számítógép interakcióját, hamar a jog határvidékeire tévedünk. Itt aztán nagyon izgalmas filozófiai és jogelméleti problémákkal (is) találkozhatunk. A dolgozat is tele van ilyenekkel, amelyek közül sokat csak felvetek, de megoldani nem tudok. Meddig tarthat a gépek terjeszkedése a jog világában? Miben különbözik a jogi szabály a programozás szabályaitól? Milyen struktúrák algoritmizálhatóak a jogban, és melyek nem? Mi a természete a jogi szövegnek, a jogi dokumentumnak? Hogyan reprezentálható a jogi tudás

számítógéppel? Hol tud a gép a jogász segítségére lenni, és hol nem? És így tovább.

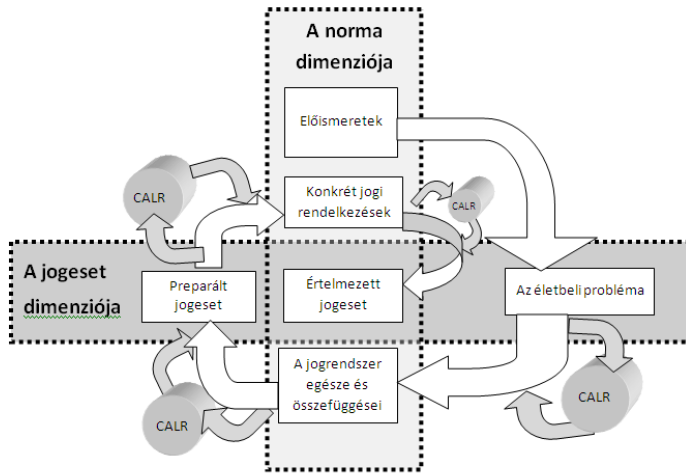
Összefoglalva, a kutatási feladat a számítógéppel támogatott jogi forráskutatás elméleti vizsgálata volt.

II. Az elvégzett vizsgálatok, elemzések rövid leírása, a feldolgozás módszerei

Az *első fejezetben* három elméletet veszek kölcsön azért, hogy a jogi forráskutatást elhelyezzem egy tágabb horizonton. Az első a háború utáni német filozófia legnagyobb alakja, H. G. Gadamer hermeneutikai filozófiája, és ezen belül is Arisztotelész interpretációja. Szerzőnk itt háromféle tudásfajtát különböztet meg, a tudományost (*episztémé*), a mesteremberit (*tekhné*), és az erkölcsi-gyakorlatit (*episztémé*). Bár Gadamer a jogi tudást ez utóbbiba sorolja, én mégis azt gondolom, (és ennek a számítógép és a jog szemszögéből nagy jelentősége van), hogy az ma már inkább a három *keveredése*: nemcsak a társas térben zajló, és erkölcsi, gyakorlati tételt bíró nyelvi erőfeszítés, (phronézisz), hanem *tekhné* (egyszerű mesteremberi tevékenység, *technológia*) is, és számos pontján támaszkodik a tudományos tudásra. A jog mindhárom mozzanatában támogatható számítógéppel, de eltérő módszerekkel. A nagy szöveges adatbázisok a természetes nyelvbe ágyazott phronézisz (korlátozott) támogatására jöttek létre.

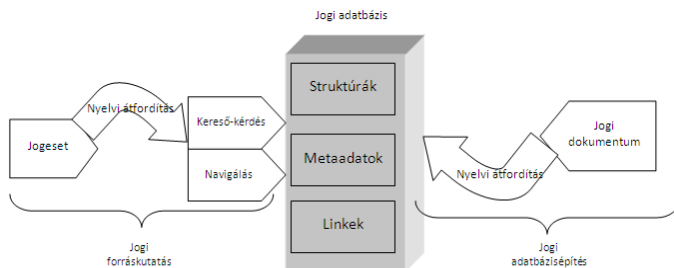
A második elmélet, melyet fejtegetéseimhez használtam, és amely részben Gadamernél is jelen van, de a német jogdogmatikában, és főként Fikentscher munkáiban

bontakozik ki, a *hermeneutikai kör* elmélete. Eszerint bármilyen szövegértelmezési feladatot körkörösén végzünk, egy értelem-elvárással, egy előismert-halmazzal fordulunk a szövegek felé, majd a részt az egészre vetítve értelmezzük, és az egészről alkotott képünket a részek fényében finomítjuk. A jogban ez azért speciális, mert itt nemcsak a szöveg egészének és részeinek egymásra vonatkoztatott értelméről, hanem a jogeset és a norma interakciójáról (sőt, mint később láthatjuk, még további tényezők összejátszásáról is) szó van. A jogi értelmezési feladat során a jogesetet a norma szóhasználatával, annak kategóriáival ragadjuk meg, konstruáljuk, a normát pedig a jogesetre tekintettel konkretizáljuk. Hogy ezt meg tudjuk tenni, rendelkezniünk kell egy előismeret-készlettel, és ezen belül egy jogival is, amelyet jellemzően az egyetemeken tanítanak. Az előismeret a jogról való gondolkodásunk legtágabb fogalmaitól egészen a jogterületekre és jogágakra vonatkozó elképzeléseinken át a részletekig, a jogintézmények tartalmáig terjed. Ennek jelentősége végigvonul a dolgozaton, mert főként ez az, ami megkülönböztet bennünket a gépektől. Ezzel az előismeret-halmazzal fordulunk a szöveges adatbázisok felé is, és ez az előismeret halmaz van ezekbe a szöveges adatbázisokba, sajátos eszközökkel belezárva. A jogi hermeneutikai, értelmezési körmozgásnak egyik részmozzanata a releváns, odaillő, felhasználható szövegek megtalálása a gépek segítségével, azaz a számítógéppel támogatott jogi forráskutatás, a *computer assisted legal research* (CALR), a dolgozat fő témája.



A harmadik elmélet voltaképp arra ad választ, hogy hogyan halad előre ez a körfolyamat. Szabó Miklós szövegfordítási teóriája ez, amely szerint a jog szövegátfordítási tevékenységek sorozataként is felfogható. Esetünkben a jogesetből előbb egy nagyjából jogi fogalmakkal megragadott eset lesz, ezután kereső-kifejezések képződnek belőle, (ezeket láthatatlanul a háttérben a gép lefordítja a maga gépi nyelvére, indexeiben kikeresi, majd találati listákat készít belőle, stb.), majd a megtalált szövegek felbukkannak, hogy egy újabb nyelvi fordítás révén beépüljenek újabb szövegekbe, beadványokba, ítéletekbe, szóbeli nyilatkozatokba. Az adatbázisok csak akkor képesek ezt a nyelvi erőfeszítést támogatni, ha bennük a szöveg maga is feldolgozott állapotban van. Ez azt jelenti, hogy az adatbázisok létrehozásakor szintén szövegfordításokat kell végezni, (metaszövegeket, metaadatokat, linkeket, szövegdesztillátumokat, indexeket, struktúrákat kell alkotni), - az

emberi tudást reprezentáló preparált szövegeket kell bennük létrehozni.



A második fejezetben a jogi adatbázisok kultúrtörténetével foglalkozom. A számítógép a jogi forráskutatás világában először Amerikában tűnt fel. Nemcsak azért, mert a heterogén és nagy mennyiségű anyag itt kiáltott ezért az eszközért, hanem azért is, mert az amerikai jogot addigra olyan módon strukturálták, hogy azt nagyon egyszerű volt számítógépesíteni. Bebizonyítom, hogy az amerikai jog a kereskedelmi kiadók jóvoltából már jóval azelőtt adatbázis-szerűen volt szervezve, az ítéletek pedig nagyüzemi-tömegtermelési módszerekkel feldolgozva, mint ahogy az elektronikus adatbázisok ténylegesen megjelentek. Ez a szerkezet rendkívül alkalmassá tette az amerikai jogot a számítógépesítésre.

Amerikában ugyanis az ipari forradalom megteremtí a valódi, papíron rögzített jogi *adatbázisokat*, főként a West kiadóvállalat jóvoltából, és olyan jogtudósok elméleteivel a háttérben mint *Langdell*, a harvardi dékán, aki a jog laboratóriumának hívja a könyvtárat, és először használja a *legal research* (jogi kutatás) kifejezést. A számítógép már szinte csak beleül a kész helyzetbe.

A kontinens az ipari forradalom hatására más irányt vesz. A kontinentális jogokban egyrészt kisebb az autoritatív textusok mennyisége, de ezt az anyagot *másképp is rendezik*, mint Amerikában, tovább folytatva a bibliamagyarázatot példaként maga elé állító *glosszátori* hagyományt. Itt a jogi *corpusban* a vertikális és horizontális kohéziót olyan struktúrák biztosították, mint a jogág, a jogintézmény, és a kódexek könyv-szerű felépítése, amelyek elméleti, filozófiai, didaktikai céllal alkotott és emiatt diffúz fogalmak és közvetítések. A kontinens jogrendszerei értelmezések és absztrakciók olyan puha hálózatán nyugszanak, - ilyet dolgoz ki szándékai ellenére a zárt logikusságra törő *Begriffjurisprudenz* (fogalmi jogtudomány) is, - amelyek reprezentálása matematikai eszközökkel nem lehetséges.

A második fejezet meséli el azt az izgalmas történetet is, ahogyan Amerikában a hatvanas években elkezdik jogi szövegek tárolására használni a számítógépet és ezzel egy időben azt, ahogy kibontakozik a *jurimetrics* mozgalom, amely nem egyszerűen ennek elméleti-tudományos háttere kíván lenni, hanem egy új jogtudomány és a tudományos, kiszámítható jog megteremtője is. A *jurimetrics* naiv programja megbukik, a jogi szövegtárolás azonban fényes sikert arat - üzletileg is.

És ennek a történetnek az egyik leágazása az, ahogy a *legal research*, a jogi forráskutatás a kiadók nyomására, Langdell dékán elméleti alapozásával, mint tantárgy és diszciplína feltűnik a 20. század elején az amerikai jogi egyetemeken. A számítógépek megjelenése után ez a tárgy aztán radikálisan átalakul, és egyre inkább, majd végül szinte kizárólag csak *computer assisted legal research*-ről (CALR – számítógéppel támogatott jogi forráskutatás) szól.

A harmadik fejezet az adatbázisokkal, ezen belül a szöveges adatbázisokkal, és ezek egyik speciális változatával a

jogi szövegeket tároló adatbázisokkal foglalkozik. Itt abból indulok itt ki, hogy a jogi munka 'phronézisz' aspektusait, a szövegekkel való munkálkodást, a 'kognitív verseny' megvívásához szükséges anyagok kikeresését és értelmezését ezek a szöveges jogi adatbázisok nagyban segítik, de egyben maguk is okozói az egyre fokozódó versenynek.

Ennek a fejezetnek az elején tisztázok néhány alapfogalmat, és röviden rámutatok, hogy a gép jelazonosítási és feldolgozási mechanizmusa radikálisan más, mint az emberi megértés. A gép ugyanis a jeleket a fizikai jegyhez társított *szabályok* révén *azonosítja*, míg az ember a jelhez *jelentést* kapcsol, és *megérti* azt.

Ennek a következményei sokrétűek. Először is, ahhoz, hogy a jogi adatbázisokkal való munka lehetséges legyen, azok hatékony segítséget nyújtsanak a jogász tevékenységében, *előtte* bennük *szövegmanipulálási előmunkálatokat, szövegátfordítási tevékenységeket* kell végezni. A szöveget darabokra kell bontani a gépnek, a megbúvó struktúrákat explicitté kell tenni, az inkonzisztenseket ki kell iktatni, indexeket, metaadatokat és linkeket kell létrehozni - a szövegeket elő kell készíteni részben a gép számára, részben meg azért, hogy aztán az ember a tudást belezárhassa ezekben az adatbázisokba. Másrészt a gépnek *úgy* kell a szövegekkel műveleteket végeznie, hogy ezek az ember számára újra fogyaszthatóak legyenek: keresőkkel és más eszközökkel kell lehetővé tenni, hogy a szabályok formájában bezárt alvó tudás eleve név váljon az emberi megértés, a jogi tevékenységek folyamatában.

A 3.3. rész az adatbázisok minket érdeklő válfajáról, a szöveges adatbázisokról beszél. Elmondja miben más a szöveges adatbázis, mint az ún. relációs adatbázis, amely a valóságot zárt értékkeszletű, sőt ha lehet számszerű deskriptorokkal akarja reprezentálni, és mitől más, mint a

nyomtatott adathalmazok. A fejezet ismerteti hogyan épülnek fel ezek az adatbázisok, és mi tartja össze őket, mitől válnak hatékony eszközökké azoknak a kezében, akik használják.

A 3.4. rész a korábbi részek megállapításaira építve a jogi adatbázisokat mutatja be. Beszél a jogi adatbázisok kitüntetett építőköveiről, a jogi dokumentumokról és ennek típusairól, valamint arról, hogy ezek mennyiben mások, mint az egyéb szöveges adatbázisok dokumentumai. Szó van itt még a speciális jogi metaadatokról és linkekről, és a jogi keresők jellegzetességeiről is, (némi összehasonlítást téve, és magyarázva az általános keresőkhöz képesti különbségeket), valamint a jogi adatbázisok felhasználói felületeiről, (amelyeken keresztül a gép és az ember érintkezik) és ezek típusairól is.

A negyedik fejezet, a számítógéppel támogatott jogi forráskutatást (CALR) taglalja, építve az előző fejezetekben leírtakra.

A fejezet a CALR-t abból a termékeny szemszögből nézi, hogy az a problémamegoldó – szöveg-előállító, és ehhez információkat kereső emberi tevékenységek nagyobb családjába (is) beletartozik. Először arra az egyszerű kérdésre akar válaszolni, hogy a *jogi probléma* mennyiben más egy számítógépes keresés szemszögeből, mint akármilyen más probléma. Azt firtatja előbb tehát, hogy mennyiben más a *jogi információ-igény*, (*information need*) mint más információ-igények, legyenek azok tudományos, vagy hétköznapi problémák, és ezen keresztül mennyiben igényelnek másfajta bánásmódot, adatkezelést, rendszereket, keresőket. Ennek kapcsán újra csak a tudás típusaihoz, a hermeneutikai kör egy speciális változatához, a *többpólusú jogi hermeneutikai erőter gondolatához*, és a szövegfordítási elmélet egy, a CALR-re adaptált speciális változatához jutunk el.

A negyedik fejezet utolsó részében a jogi kutatást mint egymást követő gyakorlati tevékenységek sorozatát is szemügyre vehetjük. Három részfázisra bontom a folyamatot, a probléma formulázásának mozzanatára, amikor az életbeli történést egy nyelvi átfordítás segítségével jogi kifejezésekre kell bontani. Ennek kapcsán megkülönböztetem a könnyű és a nehéz eseteket, és számba veszem a két eset-típus következményeit. A második fázisban, a szövegek kinyerésének fázisában a keresés és a navigáció, mint két fő módszer elemzésével próbálok a keresők működéséhez az ember-gép interakcióhoz hozzáférni, végül röviden a megtalált szövegek értékelésének folyamatát elemzem.

III. A tudományos eredmények rövid összefoglalása, azok hasznosítása illetve a hasznosítás lehetőségei

1. A jogi forráskutatásról, és különösen a géppel támogatotról, *nincsen magyar nyelven semmilyen közlemény*. Az adatbázisokat, és a jogi adatbázisokat is a legtöbb jogász a programozók belügyének tekinti, pedig ezek építése inkább jogász feladat, mint mérnöki, arról nem beszélve, hogy ezek használata már annyira elterjedt, hogy belső működésükről nem árt, ha legalább vázlatos fogalmaink vannak.

2. A gépek és az ember interakciója, és az, hogy az életünket hogyan alakítja át az Internet és a számítógép, még a hétköznapi beszélgetések szintjén is gyakran felmerülő, mondhatni életbevágóan fontos kérdés. Amennyire népszerű a számítógépek jogi *rendelkezésekre* gyakorolt hatásának elemzése, annyira ignorált *a gépeknek a jog alkalmazására, működtetésére, a jogi munka világára gyakorolt hatásainak*

végiggondolása. Jóllehet a dolgozat ennek a kérdéskörnek csak egy kis szeletével, a szöveges jogi adatbázisokkal foglalkozik, számos ponton felveti a jog és a számítógép interakciójának további problémáit is, amelyek remek kutatási témák lehetnek a jövőben

3. Mindezeket a témákat ráadásul egy *erős elméleti-filozófiai* háttér előtt teszi. Bár mind a gadameri filozófia, mind a hermeneutikai kör régóta a jogelméleti diskurzus részei, senki nem tette még fel a kérdést, hogy ezek használhatóak-e, és hogyan a számítógép és a jogász interakciójának elemzéséhez.

4. Újszerű a jogi probléma *'információ-igényként'* való ábrázolása és vizsgálata és ennek elhelyezése a többi információs igény között, elemzése a (számítógépes) keresés hatékonyságnak kontextusában.

5. Végül, még ebben a formában sehol nem kifejtett, új tudományos eredményként értékelem a *jogi hermeneutikai erőter* gondolatát, amely szerint minden jogban zajló hermeneutikai erőfeszítés (a hermeneutikai kör) egy öt pólusra bontható erőterben zajlik. Ezek a következők: (1) Az előismeretek (2) az értelmezőnek a jogi folyamatban játszott szerepe, (3) az ügy elvárt megoldásának iránya, (4) az ügyben résztvevő többi szereplő cselekedetei és nyilatkozatai, (5) végül az autoritatív szövegek, amelyek reprodukálásában a gép segítségünkre van. Ez az egyszerű értelmezési keret nemcsak a CALR tematikát helyezi el egy sokkal világosabb összefüggésrendszerbe annál, mintha csak szövegekről és ezek gépi megjelenítéséről beszélnénk, hanem általában véve az egész jogi értelmezés kérdéskörének ad egy inspirálóbb és termékenyebb elméleti háttérrel, mert nem elégszik meg a *'nyitott szövedékű/értelmezésre szoruló szöveg'* és/vagy a *'nehéz eset'* dualizmusával mint magyarázattal.

A dolgozat eredményeinek elsősorban közvetett hasznosítása lehetséges, amennyiben az informatikusok a jogot, a jogászok az informatikát érthetik meg belőle jobban. A jogelmélet képviselői talán kapnak egy újszerű látószöveget, a gyakorlati szakemberek pedig egy inspiráló gondolkodási keretet. A jogi forráskutatást hosszabb ideje tervezik a magyar jogi egyetemek tanrendjébe iktatni. Ez a dolgozat a tárgy egyik blokkjának jelentheti a megalapozását.

IV. A munka tárgykörében készült publikációk

1. Zódi Zsolt: A számítógép és a jog kapcsolata – jogelméleti alapvetés, in: Sárközy Tamás – Pázmándi Kinga (szerk.): *Az információs társadalom és a jog*, MTA Állam és Jogtudományi Intézete, Budapest, 2002.
2. Zódi Zsolt: Az információs társadalom és a jog, *Gazdaság és Jog*, 2002/7-8.
3. Zódi Zsolt: Mi kell (még) az elektronikus cégeljáráshoz? *Gazdaság és Jog*, 2007/9.
4. Zódi Zsolt: Search Engines and Legal Research, *Studia Iuridica Auctoritate Universitatis Pecs Publicata* 144 (2010)
5. Zódi Zsolt: A számítógép és a jog interakciója - a kezdetek; a Jurimetrics mozgalom az Egyesült Államokban, *Pro publico bono online extra* 2011/1.