
Zódi Zsolt,

a Budapesti Corvinus Egyetem
egyetemi docense, az MTA TK
Jogtudományi Intézet tudomá-
nyos munkatársa

E-mail: Zodi.Zsolt@tk.mta.hu

**A bírósági ítéletek
adatbázisa mint Big Data**

A „Big Data” szóról általában az intelligens hálózatokhoz kapcsolódó eszközök által generált, és a közösségi médiában keletkezett adatok jutnak eszünkbe, valamint a nevezetes (és elcsépett) „3/4/5V-definíció”¹. Így nem tűnik triviálisnak, hogy az állami szervezetrendszer termelte, legtöbbször „open data”-ként (nyílt hozzáférésű adatként) is emlegetett különféle adatok szintén ebbe a körbe sorolhatók (*NITRD* [2016]), hiszen ugyanúgy rendkívül sokféle forrásból származnak, különféle formátumúak, hatalmas mennyiségben termelődnek, és további komoly feldolgozásra, értelmezésre szorulnak ahhoz, hogy bármire is használni lehessen őket, ami az eredeti, keletkezéskori céljuktól eltér. Ezek között található relációs adatbázisokból származó adatokat épp úgy, mint szöveges állományokat.

A következőkben röviden ismertetem egy 2012-ben lefolytatott kutatásomat, amely a nyilvános bírósági ítéleti adatbázist, mint szövegkorpuszt használta kiindulópontként (*Zódi* [2014]). Az ismertetést megpróbálom tágabb kontextusba helyezni, és részben szólni a hasonló Big Data-alapú, „szövegempirikus” kutatások helyéről és szerepéről a jog területén, részben pedig néhány tanulságot megfogalmazni.

A kutatás előzményei

Mint ismert, az angolszász jogrendszerek ún. precedensjogok, amelyekben bizonyos jogterületek szabályait a bírósági ítéletek tartalmazzák, mondják ki, és bizonyos

¹ V3: méret (volume), gyorsaság (velocity), változatosság, sokoldalúság (variety), V4: érték (value), V5: igazságérték, megbízhatóság (veracity).

körben a korábbi ítéletek figyelembe vétele kötelező is a bíróságokra. Emiatt a bírósági ítéletek széles körben és gyakran hivatkoznak egymásra. A gyakorló jogászok számára már a XIX. század végén nagyon kifinomult eszközök álltak rendelkezésre a hivatkozások rögzítésére, nyomon követésére (*Bast* [1998], *Zódi* [2012]) és ez alapján az egyes ítéletek fontosságának mérésére. Ezek később a tudományometriát is befolyásolták (*Zódi* [2011], *Garfield* [1955]).

Nem csoda hát, hogy az angolszász területen már az 1950-es években megjelentek az első hivatkozási mintázatokkal, statisztikákkal foglalkozó írások, amelyek eleinte kézi számolást alkalmaztak (*Marrymen* [1954], *Merryman* [1977], *Landes–Posner* [1976], *Johnson* [1985]). Ezek a kutatások nélkülözték a bonyolult matematikai módszertani apparátust, és általában néhány nagyon konkrét kérdésre keresték a választ, mint például, hogy melyek a leggyakrabban idézett ítélettípusok, fórumok, illetve, hogy milyen lassan „amortizálódnak”, avulnak el a korábbi ítéletek stb.

Az első, a kontinentális bírói gyakorlat hivatkozásait vizsgáló kutatást (*Wagner–Döbler–Philipps* [1992]) az 1990-es évek elején végezték, már számítógép segítségével. Az általam végzett vizsgálat nagyban hasonlított ehhez, de felhasználta az azóta hatalmas számban megjelent újabb hivatkozásstatisztikai kutatások eredményeit is (*Zódi* [2015] 147–148. old.).

A céloom annak a közkeletű vélekedésnek az igazolása vagy cáfolata volt, hogy a magyar jog „precedensjogi” jellege erősödik, a bíróságokon egyre inkább szükséges nemcsak jogszabályokra, hanem korábbi felsőbbbírósági döntésekre is hivatkozni, és ez egyfajta közeledést jelent az angolszász jogrendszerek felé. Mivel ez a kérdés igen egyszerű volt, a kutatás előkészítéséhez és lefolytatásához sem használtam komplikált matematikai modelleket. A következő kérdéseket tettem fel előzetesen magamnak: 1. Milyen gyakran idéznek a bírósági ítéletek korábbi ítéleteket? 2. Vannak-e különbségek az egyes *a)* bírósági szintek, *b)* ügyszakok, *c)* ügytípusok között? 3. Milyen korábbi bírósági dokumentumokat és dokumentumfajtákat idéznek a bíróságok? 4. Változik-e (emelkedik-e) időben a hivatkozások száma? 5. Milyen tipikus szövegkörnyezetben találhatóak a hivatkozások (van-e érdemi, hosszú elemzése a korábbi ítéletnek), és kimutathatók-e ebből olyan tényezők, amelyek befolyásolják az idézést? (Például, hogy a peres felek „erőltetik” a korábbi jogesetet, vagy a pertárgy-érték befolyásolja az idézést.)

A kutatás szövegtörzsének

A kutatás törzsének – nyílt hozzáférésű adatként – az akkor már hatodik éve hozzáférhető Bírósági Határozatok Gyűjteménye volt, amely a vizsgálat időpontjában

61 512 ítéletet tartalmazott. A dokumentumok itt rtf kiterjesztésű Rich Text formátumban találhatóak. (Jelenleg a <http://birosag.hu/ugyfelkapcsolati-portal/anonim-hatarozatok-tara> oldalon). A kutatáshoz egyébként végül nem ezt az oldalt használtam, hanem ennek egy kereskedelmi célú rendszerben (az Opten Kft. OptiJUS rendszerében) található változatát, legfőképp amiatt, mert az ítélethivatkozások (az adatbázisban a linkek kiindulópontjai) viszonylag pontos felismerésére volt szükségem, amely a nyilvános adatbázisban nem volt. Tovább nehezítette a helyzetemet, hogy ez az adatbázis csak a nyilvános felületen keresztül volt lekérdezhető, amely a lekérdezésekhez használt paraméterek számát és kombinálhatóságát korlátozta. Szükségem volt ugyanis a hivatkozó és a hivatkozott ítéletek bizonyos leíró (meta-) adataira is (a hivatkozó ítéletek esetén: bíróság, ügyszak, ítélet éve, pertípus, a hivatkozottak esetén: azonosítószám, dokumentumtípus, évszám). Ezen felül, ahogy majd később ismertetem, az ítéletekből a „per/vád tárgya” információt is kinyertem, amelyhez amúgy is le kellett volna tölteni az összes dokumentumot.

1. táblázat

*A vizsgált korpusz dokumentumainak száma**
(darab)

Pertípus	Kúria	Ítéletábla	Törvényszék	Járásbíróság	<i>Összesen</i>
Polgári	4 873	9 907	10 705	1 396	26 881
Gazdasági	867	3 548	4 297	314	9 026
Munka	1 465	0	1 588	0	3 053
Közigazgatási	5 496	1 172	7 934	0	14 602
Büntető	1 166	2 704	3 066	1 014	7 950
<i>Összesen</i>	<i>13 867</i>	<i>17 331</i>	<i>27 590</i>	<i>2 724</i>	<i>61 512</i>

* Itt és a továbbiakban az adatbázis lekérdezései a 2013. év eleji állapotot tükrözik.

E korpuszban azokat a hivatkozásokat vizsgáltam, amelyek korábbi szerkesztett jogesetekre hivatkoztak.

A hivatkozott (szerkesztett) „precedens”-halmaz (a továbbiakban minden pontatlansága ellenére a precedens kifejezést fogom használni a hivatkozott dokumentumok halmazára) tartalmazta a Kúria, (régebben a Legfelsőbb Bíróság) ún. irányítási aktusait (a kifejezetten iránymutatási céllal született jogegységi határozatokat és ennek elődjét az Elvi Döntést (PED [polgári jogi tárgyú döntés], BED [büntetőjogi tárgyú döntés]), a különböző kollégiumi véleményeket és ennek elődjét a (PK [polgári jogi tárgyú], BK [büntetőjogi tárgyú], MK [munkajogi tárgyú], GK [gazdasági jogi tárgyú]) kollégiumi állásfoglalást, valamint az egyedi esetek kapcsán született,

de elvi iránymutatást tartalmazó ún. fejrésszel ellátott EBH-t [elvi bírósági határozat], és a Kúriai Döntések című jogesetközlő lapban megjelent ún. BH-kat [bírósági határozat]). A kutatásba bevontam két magánkiadású jogesetközlő lap (a KGD [Közigazgatási-Gazdasági Döntvénytár] és a BDT [Bírósági Döntések Tára], mindkettő a Wolters Kluwer Kft. kiadásában.) döntéseire történő hivatkozásokat, valamint az AB (Alkotmánybíróság) határozatait is.

A hivatkozható összes dokumentum száma 36 709 darab volt. Természetesen nem mindegyik dokumentumra történt hivatkozás, főként az 1990 előtt közzétett BH-kra szinte alig találtunk.

2. táblázat

A hivatkozott (hivatkozható) dokumentumok (precedensek)

Dokumentum (precedens)	Dokumentumok száma (darab)
Polgári és büntetőjogi tárgyú döntés és jogegységi határozat	169
Polgári jogi, büntetőjogi, munkajogi, gazdasági jogi tárgyú kollégiumi állásfoglalás és vélemény	1 162
Elvi bírósági határozat és elvi határozat	2 441
„Bírósági Határozatok” című folyóiratban közzétett döntés	20 118
„Bírósági Döntések Tára” című folyóiratban közzétett döntés	2 722
„Közigazgatási-Gazdasági Döntvénytár” című folyóiratban közzétett döntés	5 636
Alkotmánybírósági határozat	4 461
<i>Összesen</i>	<i>36 709</i>

A kutatás eredményei

A kutatás eredményei közül csak két táblázatot és egy diagramot közlök itt, a többi megtalálható a már idézett tanulmányban (Zódi [2014]).

A táblázatokból látható, hogy a magyar rendes bírósági ítéletek egyharmada tartalmaz hivatkozást precedensre. Az egyes ügyszakok hatalmas eltéréseket mutatnak. A munkaügyi messze a legtöbbit, míg a közigazgatási a legkevesebbet idéző ügyszak. E táblázatokból nem olvasható ki, de eltérés van a bírósági szintek idézési hajlandóságában is, és érdekes módon nem a Kúria, hanem az ítéltáblák a legaktívabbak, amelyek szívesen idézték a két magánkiadású, nagyjából a saját jogeseiteket közlő orgánumokat (a BDT-t és a KGD-t). Rendkívül változatosak a prece-

denstípusok idézési intenzitása is. Igaz, hogy mennyiségében a BH vezet, de ha azt is nézzük, hogy az egyes kategóriákban mennyi az idézhető dokumentumok száma, akkor messze az ún. absztrakt eszközök (a jogegységi határozat és a kollégiumi vélemény) vezetnek. A kutatás egyik fontos megállapítása az, hogy a hivatkozásokat tartalmazó dokumentumok száma az évek folyamán folyamatosan emelkedett, és 2012-re a táblabíróságokon (a legintenzívebben idéző fórumszint) már bőven 40 százalék felett volt.

3. táblázat

*A dokumentumtípusok száma bírósági ügyszakonként és impakt faktora
(darab)*

Pertípus	JEH	KV	EBH	BH	BDT	KGD	AB	A hivatkozást tartalmazó dokumentumok száma	Egyedi határozatok száma
Polgári	326	4 370	1 449	4 727	1 440	88	1 734	9 868	26 881
Gazdasági	314	812	728	1 565	799	42	75	2 685	9 026
Munka	50	502	323	1 043	12	8	119	1 634	3 053
Közigazgatási	267	1 928	347	926	19	393	517	3 604	14 602
Büntető	564	600	142	1 251	9	4	114	2 202	7 950
<i>Összesen</i>	<i>1 521</i>	<i>8 212</i>	<i>2 989</i>	<i>9 512</i>	<i>2 279</i>	<i>535</i>	<i>2 559</i>	<i>19 993</i>	<i>61 512</i>
Fajlagos idézettség (impakt faktor)	9,00	7,07	1,22	0,47	0,84	0,09	0,57		

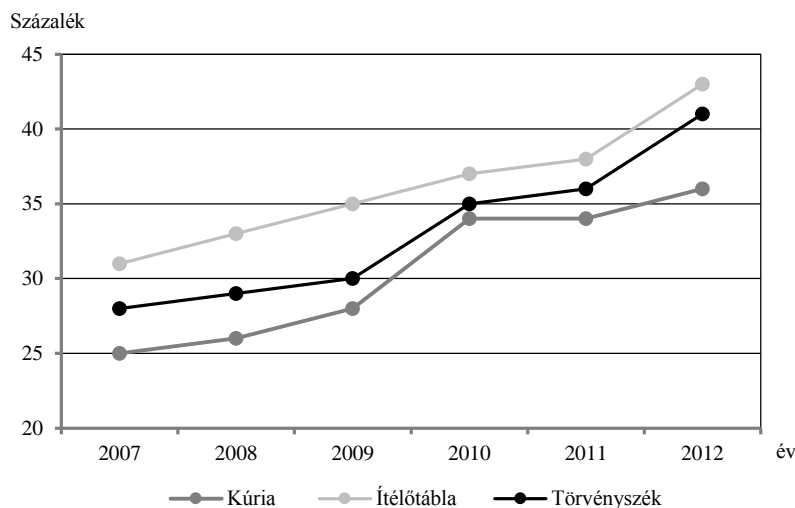
Megjegyzés. Itt és a további táblázatoknál JEH: jogegységi határozat, KV: kollégiumi vélemény, EBH: elvi bírósági határozat, BH: bírósági határozat, BDT: Bírósági Döntések Tára, KGD: Közigazgatási-Gazdasági Döntvénytár, AB: Alkotmánybíróság.

4. táblázat

*Az egyedi határozatokhoz viszonyított dokumentumtípusok aránya bírósági ügyszakonként
(százalék)*

Pertípus	JEH	KV	EBH	BH	BDT	KGD	AB	Összes
Polgári	1	16	5	18	5	0	6	37
Gazdasági	3	9	8	17	9	0	1	30
Munka	2	16	11	34	0	0	4	54
Közigazgatási	2	13	2	6	0	3	4	25
Büntető	7	8	2	16	0	0	1	28
Összes	2	13	5	15	4	1	4	33

A hivatkozást tartalmazó dokumentumok aránya bírósági szervezetek szerint, 2007–2012



Megjegyzés. A bírósági szervezetrendszerben 2012-től megváltoztak az elnevezések. Mivel a helyi bíróságok a járásbíróságoknak, a megyei bíróságok a törvényszékeknek, a Legfelsőbb Bíróság pedig a Kúriának feleltethető meg egyértelműen, mindenhol az új elnevezéseket használom. A rendszer a Kúria és az ítéltáblák jogerős, érdemi döntéseit és ezek előzményeit tartalmazza. Azaz járásbírósági döntés abba csak igen csekély számban kerül, lényegében csak akkor, ha egy jogerős törvényszéki döntést felülvizsgálatnak a Kúrián. A kis elemszám torzításokat eredményezne, ezért a járásbíróságok ábrázolásától eltekintek.

A kvalitatív kutatás és ennek tanulságai

A kutatásban kiválasztottunk egy reprezentatív ítélethalmazt, amelyet elolvastunk, és a benne található információkat rögzítettük, részben amiatt, hogy kontrolláljuk a gépi lekérdezés számait, részben pedig azért, hogy – ahogy említettem – megnézzük az idézések szövegkörnyezetét is. Rendkívül tanulságosnak találom, hogy az igen sok adattal és következtetéssel szolgáló gépi statisztikák mellett a kutatás csaknem legfontosabb következtetéseit mégis ennek a kvalitatív résznek köszönhetjük.

Már a gépi lekérdezés során világossá vált, hogy pusztán az üzyszakokkal nehezen magyarázhatók az idézési hajlandóság hatalmas eltérései, hiszen például a gazdasági üzyszak és a polgári üzyszak csaknem ugyanazt a jogi rendelkezéshalmazt (nagyraérett a Ptk.-t) alkalmazza, csak az előbbi gazdálkodó szervezetek konfliktusai-ban, míg az utóbbi magánszemélyek pereiben. Sőt, a két üzyszak közötti különbségek (például a kötelező perbeli képviselő a gazdasági ügyekben – az ügyvédek nyil-

ván szívesebben hivatkoznak precedensekre, illetve jellemzően magasabb a per-tárgyérték, amely esetleg több jogforrás, precedensen alapuló érv bevonását indokolja) épp fordított helyzetet kellett volna mutassanak. A gazdasági ügyszakban ugyanis mintegy 7 százalékkal kisebb volt a hivatkozó dokumentumok száma.

A kiválasztott ítéletek elolvasása és kódolása során egyértelművé vált, hogy a magyarázat a szűkebb pertípus (percsoport, téma) hatásában keresendő, ez ugyanis igen erősen befolyásolja az idézési hajlandóságot. Ezért előbb géppel, majd kézi tisztítással megpróbáltam a perek tárgyát („témáját”, a büntető ügyekben a „vád tárgyát”) kiemelni a szövegekből. Ez csak bizonyos szövegfordulatok gépi elemzésével volt lehetséges, valamint a polgári, gazdasági, közigazgatási és munkaügyi ügyszakokban ráadásul több, mint 11 ezer féle kombinációt eredményezett, amit ezután megpróbáltam nagyobb felismerhető csoportokba sorolni.

Miután ez sikerült, kiderült, hogy a hivatkozások sűrűségét valóban csaknem kizárólag a per/vád tárgya befolyásolja. Vannak olyan pertípusok (például a sajtóhelyreigazítás vagy személyiségi perek), ahol szinte minden ítélet tartalmaz precedenshivatkozást, és ezek a típusok épp a polgári ügyszakban vannak.

A másik lényeges következtetés szintén az elolvasás nyomán rajzolódott ki. Magyarországon vitatott, hogy valóban a precedensrendszer felé haladunk-e, a szakirodalom is megosztott a témában. A kutatás kvalitatív, kézi részéből kiderült, hogy valójában mindkét tábornak igaza van: a hivatkozások száma valóban dinamikusan emelkedik, és ez alapján látszólag valóban haladunk a precedensrendszer felé, azonban a hivatkozások szövegekörnyezete azt mutatja, hogy a precedenseket nem angol-szász módszerekkel kezeljük és használjuk, hanem „kontinentális módon”. Ez elsősorban azt jelenti, hogy a bíróságok az esetek elsősorban többségében kizárólag az ítélet hivatalos összefoglalóját, a fejrészt idézik, vagy még azt sem, csak a precedens számát. Így a precedensrendszerek tényállásokat összehasonlító, azok különbségét elemző módszere (az ún. tényintenzív érvelés (*Komárek* [2013])) egyáltalán nem jellemző a magyar bíróságok döntéseire.

A legfontosabb tanulság ezek után tehát az, hogy egy hagyományos „kvalitatív” kutatásnak mindig ki kell egészítenie a kvantitatív kutatási részeket, és nemcsak, sőt nem is elsősorban azért, mert a kvantitatív „tévedne”, hanem elsősorban amiatt, mert csak a kvalitatív, azaz a hagyományos fogalmi keretekben mozgó, a nyers számokat és összefüggéseket a hagyományos narratívákba visszahelyező eredményekkel tudunk egyáltalán valamit kezdeni. Ez különösen igaz az olyan Big Data-típusú, azaz nagy, többnyire strukturálatlan szöveghalmazokat gépi eszközökkel vizsgáló kutatásokra, ahol a kutatás kimenetele nagyrészt nyitott.

Az ilyen kutatásoknak a jogtudomány területén van egy további jellegzetessége. A hagyományos „dogmatikai”, azaz fogalomelemző módszert, amely csaknem ezer éven át uralta a jogtudományt, ma sokan bírálják, épp pontatlansága miatt (*Jakab* [2016] *Bódig* [2016]). Ugyanakkor a Big Data-jellegű kutatások a jog területén nem

tudnak a dogmatikai, fogalmi hálók nélkül meglenni, hiszen maguk is szövegeket elemeznek, és a szövegek értelmi struktúráját ezek a fogalmak teremtik meg.

A hagyományos (dogmatikai) jogtudománynak ezen felül még egy további, speciális jellegzetessége, hogy különleges és szoros kölcsönkapcsolata van a joggyakorlattal. A joggyakorlatot és annak fogalmait elemzi, de oda „adja vissza” az érveit használatra, így a jogtudomány és a gyakorlat között folytonos felelgetés, körforgás van. Ennek a kapcsolatnak bizonyos kapuőri szerep is fontos része. A dogmatika a joggyakorlat irányába mind az érveket, mind a konkrét személyeket (a jogi egyetemen keresztül) megszüri.

Mindez azt jelenti, attól, hogy rejtett összefüggésre bukkantunk egy szövegempirikus statisztikai elemzéssel, még elképzelhető, hogy ezt az összefüggést a dogmatikai jogtudomány egyszerűen ignorálni fogja, vagy épp irrelevánsnak tekinti. Tudomásul kell vennünk, hogy az adatalapú érvelésnek a jogi érvelés világában csak korlátozott szerepe lehet. Ugyanakkor az ilyen és hasonló szövegalapú Big Data-kutatások sok olyan tényezőre is rámutatnak, amely a jogászokról is megmutathatja, hogy „nem tudják, de teszik”.

Irodalom

- BAST, C. M. [1998]: What's in a name? Shepard's citators. *Journal of Paralegal Education and Practice*. Vol. 14. pp. 41–58.
- BÓDIG M. [2016]: A jogtudomány módszertani karaktere és a dogmatikai tudomány eszméje. In: *Bódig M. – Zódi Zs. (szerk.): A jogtudomány helye szerepe és haszna; tudomány módszertani és tudományelméleti írások*. MTA TK Jogtudományi Intézet, Opten Informatikai Kft. Budapest.
- GARFIELD, E. [1955]: Citation indexes to science: A new dimension in documentation through association of ideas. *Science*. Vol. 122. Issue 3159. pp. 108–111. <http://dx.doi.org/10.1126/science.122.3159.108>
- JAKAB A. [2016]: Jogi érvelések és dogmatikai jellegzetességek összehasonlítása számokkal – jogösszehasonlítás, matematika és jogbölcselet. In: *Hegyí Sz. – Vinnai E. – Zódi Zs. (szerk.): Orbis Iuris – Ünnepi tanulmánykötet Szabó Miklós 65. születésnapjára*. Bíbor Kiadó. Miskolc.
- JOHNSON, C. A. [1985]: Citations to authority in supreme court opinions. *Law & Policy*. Vol. 7. Issue 4. pp. 509–523. <http://dx.doi.org/10.1111/j.1467-9930.1985.tb00365.x>
- KOMÁREK, J. [2013]: Reasoning with previous decisions: Beyond the doctrine of precedent. *American Journal of Comparative Law*. Vol. 61. Issue 1. pp. 149–172. <http://dx.doi.org/10.5131/AJCL.2012.0013>
- LANDES, W. A. – POSNER, R. A. [1976]: Legal precedent: A theoretical and empirical analysis. *Journal of Law & Economy*. Vol. 19. No. 2. pp. 249–307. <http://dx.doi.org/10.1086/466868>
- MERRYMAN, J. H. [1954]: The authority of authority: What the California Supreme Court cited in 1950. *Stanford Law Review*. Vol. 6. No. 4. pp. 613–673. <http://dx.doi.org/10.2307/1226493>

- MERRYMAN, J. H. [1977]: Toward a theory of citations: An empirical study of the citation practice of the California Supreme Court in 1950, 1960, and 1970. *Southern California Law Review*. Vol. 50. pp. 381–428.
- NITRD (NETWORKING AND INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT) [2016]: *The Federal Big Data Research and Strategic Plan*. The Networking and Information Technology Research and Development Program. May. <https://www.nitrd.gov/pubs/bigdatardstrategicplan.pdf>
- WAGNER-DÖBLER, R. – PHILIPPS, L. P. [1992]: Präjudizien in der Rechtsprechung: Statistische Untersuchungen anhand der Zitierpraxis deutscher Gerichte. *Rechtstheorie*. Band 23. Heft 2. pp. 228–242.
- ZÓDI Zs. [2011]: A Google, a jogi adatbázisok és a szöveg számítógépes uralásának három módja. *Infokommunikáció és Jog*. 46. sz. 175–178. old.
- ZÓDI Zs. [2012]: *Jogi adatbázisok és jogi forráskutatás – Gépek a jogban*. Gondolat Kiadó. Budapest.
- ZÓDI Zs. [2014]: *A korábbi esetekre történő hivatkozások mintázatai a magyar bíróságok ítéleteiben*. MTA Law Working Papers. 2014/1. Magyar Tudományos Akadémia. Budapest. 1–55. old.
- ZÓDI, Zs. [2015]: Citations of previous decisions and the quality of judicial reasoning. *Acta Juridica Hungarica*. Vol. 56. Issue 2–3. pp. 129–148. <http://dx.doi.org/10.1556/026.2015.56.2-3.3>