

Jogszabályok hivatkozásainak automatikus felismerése és a belső hivatkozások struktúrája

Hamp Gábor¹, Syi¹, Markovich Réka^{2, 3}

¹ BME Szociológia és Kommunikáció Tanszék 1111 Budapest, Egy József u. 1.
hampg@eik.bme.hu, i@syi.hu

² ELTE Filozófiatudományi Doktori Isk., Logika Tanszék 1088 Budapest, Múzeum krt. 4/I

³ BME Üzleti Jogi Tanszék, 1111 Budapest, Magyar Tudósok körútja 2.
markovich@phil.elte.hu

Kivonat: A jogrendszert alkotó jogszabályok nem egymástól független létezők: hálózatot alkotnak – ebből és hierarchikus viszonyukból következően beszélhetünk rendszerről. Ebből következően a jogszabályokban sok hivatkozást találunk más jogszabályokra. De mivel a hivatkozás a másik szöveghely behívásával redundanciát kerül, nemcsak egy másik jogszabály szöveghelyére történő hivatkozásnak van értelme: az adott jogszabályon belül történő hivatkozás szintén gyakori. A jogszabályok automatikus logikai elemzésének aspektusait feltárni célzó kutatásunkban szükségessé vált a hivatkozások kezelése. Az automatikus hivatkozásfelismerés a logikai elemzés számára történő szöveg-előkészítés mellett más, igen komoly gyakorlati jelentőséggel is bír.

A jogszabályok alkotta szöveggörnyezet folyamatosan változik: új jogszabályok születnek, mások részben vagy egészében hatályukat veszítik. A jogszabályi környezet változásai miatt a hatályban lévő jogszabályok hivatkozásai könnyen és gyakran elavulnak. Ebben a helyzetben elemi gyakorlati érdek, hogy az elavult hivatkozások feltárását, frissítését gépi eszközökkel lehessen támogatni. Mi a kutatásunkban a jogszabályok logikai elemzése automatizálásának aspektusait vizsgáljuk. Tekintettel arra, hogy a logikai elemzés a szöveg mondatszintű egységein kivitelezhető, a szöveg-előkészítés részeként vált szükségessé – a gyakorlati jelentőséggel is bíró – hivatkozáskezelés.

E cél eléréséhez első lépésként a jogszabályi hivatkozások felismerését, valamint a teljes szövegen belüli pozícióinak gépi azonosítását kell elvégeznünk, majd a hivatkozásból ki kell nyerni a hivatkozott jogszabály vagy a hivatkozott jogszabályi szerkezeti egység azonosításához szükséges információt. A továbbiakban – ha az egyértelmű tárgyalásmód azt megkívánja – a hivatkozott jogszabályi egységet (a jogszabály egészét vagy annak egy részét) célnak (céljogszabálynak), a hivatkozás tartalmazó jogszabályi egységet forrásnak (forrásjogszabálynak) nevezzük. A forrásoldali tehát a hivatkozó, a céloldali a hivatkozott jogszabály (jogszabályi szerkezeti egység). Ha a forrás- és a céljogszabály egybeesik, belső hivatkozásról beszélünk.

Amennyiben nem esik egybe, azaz külső hivatkozással állunk szemben, a meghatározó lépések egyike a hivatkozott jogszabály beazonosítása. A jogszabályoknak – elméletileg – egyedi címük van, így egyértelműen lehet rájuk hivatkozni. A jogszabályszerkesztésről szóló rendelet pontosan előírja, hogyan kell a jogszabályok címét megadni, és hogyan kell rájuk hivatkozni [2]. Az elfogadott hivatkozási módok közül nem mindegyik alkalmas arra, hogy teljes biztonsággal fel lehessen ismerni valamely

jogszabály címét kizárólag szintaktikai mintázat alapján, ezért a jogszabály-hivatkozások felismeréséhez szükség van egy olyan külső címtárra, amely tartalmazza az összes jogszabály összes lehetséges címét. Ennek hiányában a jogszabályokra való hivatkozások felismerése nem garantált. Egy ilyen címtár webes források feldolgozásával előállítható [10], mi is létrehoztunk egy ilyen forrást, de ebben a tanulmányban a külső linkekkel érdemben nem, csak problémafelvetés szintjén foglalkozunk.

A jogszabályszerkesztésről szóló rendelet meghatározza a jogszabályok lehetséges szerkezeti egységeit és azok hierarchiáját: alpont, pont, bekezdés, szakasz, alcím, fejezet, rész és könyv) [2]. A szerkezeti egységeket jelöléssel (címekekkel, számokkal, egyéb jelekkel) látja el a jogalkotó – az egyértelmű azonosítás és ezzel az egyértelmű hivatkozási lehetőség érdekében. A hivatkozás egy konkrét jogszabály adott szerkezeti egységén belül helyezkedik el, és vagy egy jogszabályra teljes egészében, vagy egy (akár az azonos, akár egy másik) jogszabály adott szerkezeti egységére mutat. A hivatkozások gépi felismeréséhez ismernünk kell a hivatkozások lehetséges típusait, és egyértelmű eredményeket nyújtó azonosítási technikát kell használnunk a jogszabályokra és azok szerkezeti egységeire történő utalásokhoz.

A jogszabályszerkesztésről szóló rendelet III. Fejezete elkülöníti a *rugalmas*, valamint *merev hivatkozásokat*, azaz a tartalmi körülírással megadott, illetve az adott jogszabály azonosítóit tételesen megadó utalásokat, valamint megkülönbözteti a *belső*, illetve *külső hivatkozásokat*. Utóbbi felosztás szerint arra figyelnek, hogy az adott jogszabály saját magára, pontosabban saját magán belül egy másik szerkezeti egységre vagy egy másik jogszabály egészére vagy részére utal. A rendelet megadja ezeknek a hivatkozástípusoknak a használati szabályait és szintaxisát. A gépi elemzéshez azonban további tipizálásra is szükségünk van. El kell különíteni az *egyszerű* és a *halmazott hivatkozásokat*, valamint az *egyszeres* és *többszörös hivatkozásokat* egymástól. A halmazott hivatkozások több szerkezeti egységre utalnak a céldoldalon (pl. ...a 37. § (2)-(5) bekezdésében...). Többszörös hivatkozásokról akkor beszélünk, ha az adott szerkezeti egységben egynél több egyszerű vagy halmazott hivatkozás van (pl. *Az (1) bekezdés a) pontjában felsoroltak közösen, illetve az (1) bekezdés b)-d) pontjában meghatározottak közösen is gyakorolhatják a fenntartói jogokat.*); ha egy szerkezeti egységben belül egyetlen hivatkozás van, akkor arra egyszeres hivatkozásként utalunk. Így háromféle hivatkozáshalmazással találkozhatunk. Létezik tehát

- i) forrásoldali halmazott hivatkozás, amikor a forrásoldali szerkezeti egységben több, egymástól egyértelműen elkülöníthető hivatkozás szerepel (*a 80. § (2) bekezdésének a) pontja, valamint a 83. szakasz (1) bekezdésének f) pontja*),
- ii) céldoldali halmazott hivatkozás, mikor egy egyszerű forrásoldali hivatkozás több forrásoldali pontra mutat (*a 80. § (2) bekezdésének a)-c) és f) pontjai*),
- iii) kétoldali vagy kétszeresen halmazott hivatkozás, amikor egyrészt a forrásoldalon több hivatkozás van egy szerkezeti egységben belül, másrészt a forrásoldali hivatkozások közül legalább egy többszörös céldoldali hivatkozást tartalmaz (*a 80. § (2) bekezdésének a)-c) pontjai, valamint a 83. § (1) bekezdésének f) és g) pontjai*).

Az a tény, hogy a jogszabályok jelölt szerkezeti egységekből állnak, lehetővé teszi azt, hogy a jogszabályok bármelyik szerkezeti egységére egyértelműen tudjunk hivatkozni. Figyelni kell azonban arra, hogy a szerkezeti egységekhez rendelt jelölések nem feltétlenül egyediek, hiszen míg a szakasz számozása folyamatos, addig a pontok

számozása minden szakasz alatt, az alpontok számozása minden pont alatt újrakezdődik, így önmagukban nem alkalmasak az egyértelmű azonosításra. Tehát pl. 1. §-ból csak egy lehet egy törvényben, (1) bekezdésből több is. Az egyértelmű hivatkozást azáltal biztosíthatjuk, ha kihasználjuk a jogszabályok hierarchikus tagoltságát. A hierarchiára építve úgy hivatkozhatunk egyértelműen, hogy az adott szerkezeti egység jeléhez sorban hozzákapcsoljuk az öt magába foglaló összes szerkezeti egység jelét is, vagyis adott szerkezeti egységet azzal a *hierarchikus jelsorozattal* azonosítunk, amelyet az egymásba ágyazott szerkezeti egységekhez tartozó jelekből állítunk össze.

x. évi y-ról szóló z. törvény + 2. § + (1) bekezdés + a) pont + ab) alpont

A fenti kifejezéssel egyértelműen rámutathatunk a szóban forgó szerkezeti egységre annak ellenére, hogy a) pontból is, ab) alpontból is több lehet az adott jogszabályon belül.

Mivel a jogszabályok szerkezeti egységei meghatározott típusokba sorolhatóak, egy ilyen hierarchikus jelsorozat megadható azzal, hogy az egyes összetevői milyen típusba tartoznak és az adott típusnak milyen azonosítója van. A fenti példában szereplő jelsorozatot így formalizálhatjuk:

<j:123;s:1;b:2;p:a;a:ab>

ahol a 'j' a jogszabály, az 's' a szakasz (§), a 'b' a bekezdés, a 'p' a pont, az 'a' az alpont mint szerkezetiegység-típusok jele, a kettőspont után következő szám- és betűkombinációk pedig a konkrét szerkezeti egységek – jogszabályban szereplő – jelei. Egy szerkezeti egységnek a fenti formalizmust követő azonosítója kinyerhető az adott szerkezeti egységben szereplő explicit, valamint a szerkezeti egység hierarchikus beágyazódásáról vonatkozó implicit információkból. Ha rendelkezésre áll egy jogszabálycímtár erőforrásként, akkor azt arra is használhatjuk, hogy a jogszabályok egyedi jeleként a címtárban található egyedi azonosítót használjuk 'j' értékeként. Ha így teszünk, akkor az egymásba ágyazott szerkezeti egységek típusából és jeléből álló szimbólumkettősök sorozatával minden szerkezeti egységet egyedi módon azonosíthatunk. Ez a jelsorozat azért is fontos, mert ugyanezt a mintázatot előállíthatjuk az automatikusan megtalált és felcímkézett hivatkozásokból is. Amikor megtalálunk és elkülönítünk egy hivatkozást, azt összetevőkre bontva ugyanolyan módon felépíthetünk egy jelsorozatot, ahogy azt a szerkezeti egységek azonosításakor tesszük. Az elemzett jogszabályszövegekből kinyert hivatkozásokat jellemző jelsorozat értékeit pedig összehasonlíthatjuk a jogszabálykorpuszban található céloldali szerkezeti egységek jelsorozataival, és ha azok között megtaláljuk a keresett mintázatot, akkor összekapcsolhatjuk a forrást a céloldali szerkezeti egységgel. Ez az összekapcsolás természetesen csak akkor (és olyan mértékben) működik, amikor (és amilyen mértékben) megtalálhatóak a hivatkozott jogszabályok a korpuszunkban.

A gépi feltárás célja az, hogy a hivatkozásokat megtaláljuk a forrásoldalon, feltárjuk, hogy hány hivatkozás van az adott szerkezeti egységen belül, majd minden egyes hivatkozást felbontva beazonosítsuk azokat a jogszabályokat, illetve azon belüli szerkezeti egységeket a céloldalon, amelyekre a hivatkozások mutatnak. A feladat tehát az, hogy az elemzett jogszabályszöveget belül:

- i) megtaláljuk a teljes hivatkozás kezdő- és zárópozícióját (többszörös hivatkozás esetén több értékpárt kell megtalálnunk),

- ii) feltárjuk, hogy egyszerű vagy halmozott hivatkozásról van-e szó (utóbbi esetben elkülönítjük a hivatkozási célpontok közös, illetve egyedi összetevőit),
- iii) azonosítjuk az egyes hivatkozásokban hivatkozott jogszabályt, és
- iv) meghatározzuk a hivatkozott szerkezeti egység(ek) azonosító jelsorozatát.

A hivatkozás homogén abban az értelemben, hogy mind a forrásoldalon, mind a céloldalon szerkezeti egységeket kapcsol össze. Ez a homogenitás összhangban van azzal a tézissel, amely a jogszabályok legkisebb – hivatkozási – egységének a szerkezeti egységeket tekinti (és nem engedi meg az azokon belüli, például mondatokra irányuló utalásokat). A gépi hivatkozásfelismerés gyakorlatában azonban többféle inhomogenitással is találkozhatunk. Egy hivatkozás két végpontjának pontos meghatározása eltér a jogi korpuszon belül: a céloldalon csak a szerkezeti egységet adjuk meg (az öt azonosító jelsorozattal), a forrásoldalon ellenben meg kell adni a hivatkozás szerkezeti egységen belüli pozícióját is. Utóbbi esetben pontosabb lokalizálást kell elvégeznünk. Erre azért van szükség, mert egy szerkezeti egységen belül előfordulhat többszörös hivatkozás, amiket el kell tudnunk különíteni egymástól, és meg kell tudnunk mondani, hogy melyik hivatkozás hol helyezkedik el a szerkezeti egységen belül.

A hivatkozások inhomogének lehetnek abban az értelemben is, hogy az – egyetlen – forrásoldali szerkezeti egység több szerkezeti egységre mutat a céloldalon. Akkor fordulhat elő ilyen helyzet, amikor a hivatkozás olyan szerkezeti egységre mutat, amelynek vannak alárendelt szerkezeti egységei. A forrásoldalon pl. csak annyit adnak meg, hogy a céloldalon egy másik jogszabály valamely bekezdésére hivatkoznak (ebben az értelemben ez egy egyszerű hivatkozásnak számít), ám a hivatkozott jogszabály ismeretében kiderülhet, hogy a szintaktikai értelemben véve egyszerű utalás szemantikai értelemben valójában halmozott hivatkozásnak minősíthető, hiszen több szerkezeti egység – pont és alpont – tartozik a terjedelmébe. Ezeket nevezhetjük rejtett vagy látens halmozott hivatkozásnak. Ennek legnyilvánvalóbb esete az, amikor egy hivatkozás egy másik jogszabály teljes szövegére mutat.

A céloldali hivatkozások feloldásának további nehézségét jelentik azok az – igen gyakori – esetek, amikor egy szerkezeti egység egyszerre kétféle minőséggel is rendelkezik. A bekezdések esetében tipikus megoldás az, hogy ugyanaz a szerkezeti egység a bekezdés első szakasza is egyben. Az ilyen helyzetek kétféle problémát jelentenek. Egyrészt könnyen megtörténhet az, hogy az első szakaszra utaló jelölést nem teszik ki (például azért, mert a bekezdés csak egy szakaszból áll). A szerkezeti egység felismerő modulnak képesnek kell lennie arra, hogy kezelje az ilyen helyzeteket. Másrészt könnyen zavart okozhat az is, hogy a forrásoldalon egyszer úgy hivatkoznak az adott bekezdésre, hogy a bekezdés egészére akarnak utalni, máskor viszont csak a bekezdés első szakaszára (ami „fizikailag”, szintaktikailag ugyanaz a szerkezeti egység). Ezt a kétféle hivatkozási igényt csak úgy tudjuk kielégíteni, ha az ilyen bekezdésekhez (szerkezeti egységekhez) kétféle tipizálást rendelünk, és így kétféle hivatkozási lehetőséget biztosítunk.

A jogszabályok szerkezeti egységeire való hivatkozásnak szintaktikailag jól felismerhető szabályai vannak. Ezek a hivatkozások jobbról könnyen lezárhatóak: könnyen meg lehet találni a hivatkozási sor utolsó elemét. A hivatkozás kezdetének megtalálása a külső hivatkozások esetében jóval nehezebb feladat. Elvileg mindig a tartalmazza a jogszabály „nevét”. Korábban említettük, hogy a jogszabályokra (pontosabban a jogszabályok címére) való hivatkozás lehet merev vagy rugalmas, ami a

kontextusmentes vagy a kontextusfüggő hivatkozási technikának felel meg (mások ezt a dichotómiát direkt vagy relatív hivatkozásnak nevezik [8]). A kontextusmentes cím-hivatkozás felismerése elvileg viszonylag problémamentes kellene hogy legyen, hiszen a hivatkozás a jogszabály teljes címét tartalmazza, aminek szigorú – a jogszabályszerkesztésről szóló rendelet előírásait követő – mintázatát fel lehet ismerni, ám e felismerés paraszolási technikákat igényel a jogszabály címének pontos beazonosításához. A rugalmas hivatkozás detektálása még nehezebb, hiszen a felismeréshez további információra, kontextusismeretre van szükség. A gyakorlatban minden belső hivatkozás kontextuális hivatkozásnak minősíthető, hiszen egy konkrét jogszabályon belül sosem fognak a szöveg egy másik pontjára a jogszabály teljes címével hivatkozni. Egyszerűbb eset, amikor az adott szövegben a jogszabály önmagára úgy hivatkozik, hogy „e törvényben”, hiszen ekkor (és ebből) tudhatjuk, hogy a törvény itt saját magára hivatkozik (más kérdés, hogy vajon van-e értelme ezt valódi hivatkozásként kezel-nünk). Hasonló módon ismerhetők fel azok a belső hivatkozások, amelyek alacsonyabb szintű szerkezeti egységre utalnak (pl *e bekezdés a) és b) pontja*). Bár a belső hivatkozások csak kontextuálisok lehetnek, azért ennek a kontextualitásnak lehetnek fokozatai. Vannak erős – kontextuális – hivatkozások, amikor adott hierarchikus szintre vonatkozó utalást egyértelműen adnak meg (*a 3. § (2) bekezdés a) pontja*), és vannak gyenge hivatkozások, amikor a hivatkozás valamely eleme relatív, kontextusfüggő (*e § (2) bekezdése*).

A külső hivatkozások között vannak nem kontextuális hivatkozások (*a nemzeti felsőoktatásról szóló 2011. évi CCIV. törvény 3. § (2) bekezdése*), de előfordulhatnak köztük kontextuálisok is, amelyeken belül elkülöníthetünk további altípusukat. Gyakori megoldás, hogy a jogszabály elején bevezetnek a későbbiekben többször hivatkozni kívánt külső jogszabályra valamilyen rövidítést, és a továbbiakban ezt az alakot használják a linkek felépítésekor. Ezek a rövidítések gyakran olyan alakúak, amelyek más jogszabályokban is előfordulnak, az adott jogszabály egyedi azonosítójaként, rövid címeként szolgálnak. A kontextualitás erősebb fokát jelenti az a megoldás, amikor olyan rövidítést alkalmaznak, amely csak az éppen adott jogszabályban érvényes, és semmiképpen sem vihető át jogszabályról jogszabályra. Beszédes példája ennek a *Módtv.* rövidítés, ami mindig egy módosító törvény alkalmilag bevezetett – és ezáltal – alkalmi rövidített megnevezése, azonban módosító törvényből nagyon sok van, és ez a minősége egyrészt csak az adott kontextusban érdekes, másrésztől a rövidítés jelölete ennek megfelelően változó.

Fentebb bemutattuk, hogy a hivatkozásokat és a szerkezeti egységeket ugyanazzal a logikával bonthatjuk fel komponensekre, ugyanazzal a technikával azonosíthatjuk egyedi módon. Az azonosításban kiemelt szerepe van az azonosító jelsorozat kezdő értékének, amely a szerkezeti egységet magába foglaló jogszabály egészére utal. Cél-szerű ezt a kitüntetett szerepet azzal jelezni, hogy ezt az első komponenst nevezzük a hivatkozás *fejének*. A hivatkozások tipizálásakor, a kontextusok kezelésekor a hivatkozás fejében található információt hasznosíthatjuk. Az, hogy külső vagy belső hivatkozásról van-e szó, csak akkor derül ki, amikor már megtaláltuk a hivatkozás fejét, vagyis már ismerjük a hivatkozásba foglalt jogszabály címét. Ha ez megegyezik az éppen elemzett jogszabály címével, akkor belső hivatkozásról van szó, ha eltér a két cím egymástól, akkor külső hivatkozásról beszélhetünk.

A halmozott hivatkozásoknak több típusa lehet. Előfordulhat olyan halmozás, amikor egymástól teljesen elkülöníthető, semmilyen átfedést, közös részt nem tartalmazó

egységeket sorolnak fel egymás után. Ilyen esetben mindegyik hivatkozást önmagában lehet feltárni. Azok a halmozott hivatkozások, amelyek ugyanazon jogszabály több, különböző helyére mutatnak, egyfajta hivatkozásfelsorolásnak is tekinthetők. Az ilyen felsorolásoknak több típusa is előfordulhat. Lehetnek „folytonos” hivatkozások, amikor szigorúan egymás után követő, azonos típusú szerkezeti egységek sorozatára utalnak egy kötőjeles rövidítéssel: *a 23-37. §.* Az egyértelmű céloldali lokalizálás érdekében az ilyen sorozatot fel kell bontani elemeire, hogy minden egyes szerkezeti egységre – külön-külön – rá tudjunk mutatni.

A felsorolás másik fajtája a diszkrét hivatkozási sor, amikor folytonos sorozatba nem rendezhető elemeket vesszővel elválasztva sorolnak fel: *a 16. § (2), (3), (5) bekezdése.* Ilyenkor könnyebb feladat felbontani a sorozatot elemi egységekre, de ekkor is meg kell tenni, hogy a célirányban egyértelmű utalásokhoz juthassunk.

Ez a jelenség nem keverendő össze azzal a – korábban leírt – helyzettel, amikor olyan látens halmozott hivatkozásról van szó, amely a forrásoldalon egyetlen, fel nem bontható hivatkozást találunk, ami viszont a céloldalon mégis több szerkezeti egységet fog össze, mert a hivatkozás magasabb hierarchikus szintű egységre mutat.

Az alkalmazott technika

A jogszabályok szerkezeti egységeinek tipizálására, a szerkezeti egységek közti hierarchikus alárendeltségi (szülő-gyerek) kapcsolatok feltárására, a szerkezeti egységeket egyértelműen azonosító hierarchikus jelsorozat készítésére, a Goody-listák megtalálására [5, 7], tipizálására, illetve a Goody-mondatok előállítására reguláris kifejezéseket használunk.

Speciális hivatkozások, kivételkezelés

A normál, szabályos hivatkozások mellett olykor előfordulnak speciális, atipikus utalások is a jogszabályokban. Ezek egy része – különböző okok miatt – nem is igazán tekinthető valódi hivatkozásnak. Vegyük sorba őket.

A hivatkozások között előfordulhatnak *egymásba ágyazott hivatkozások*. Ilyen esetekben egy adott jogszabály adott szerkezeti egységére való hivatkozáson belül egy másik jogszabály valamely szerkezeti egységére mutató hivatkozást találhatunk. Íme, egy példa erre: *e törvénynek a 2012. évi CCVIII. törvénnyel megállapított 85/A. § (1) bekezdése.* Az idézetet úgy kell értelmeznünk, hogy egy belső hivatkozásban (*e törvénynek a ... 85/A. § (1) bekezdése*) szerepel egy külső (*2012. évi CCVIII. törvénnyel*) hivatkozás (amely épp itt a törvény egészére, nem pedig annak egy szerkezeti egységére mutat).

Vannak jogszabályok (tipikus módon törvények), amelyekben – a jogszabály által szabályozni kívánt területhez tartalmilag egyáltalán nem kapcsolódva – a jogszabály valamelyik szerkezeti egységében megváltoztatják egy másik jogszabály valamely szerkezeti egységének a szövegét. Ezek a betétek tartalmilag függetlenek a „befogadó” szöveghez képest, ezért vendégszövegeknek tekinthetők, az ilyen hivatkozásokat ezért nevezhetjük *vendégszöveg-hivatkozásoknak*. Nézzük meg a következő példát:

(2) A Magyarország helyi önkormányzatairól szóló 2011. évi CLXXXIX. törvény 109. § (2) bekezdése a következő szöveggel lép hatályba:

„(2) A képviselő-testület kizárólag a nemzeti vagyonról szóló törvényben meghatározott személyekkel köthet vagyonkezelési szerződést.” [3]

Ez a sajátosság azért említésre méltó a hivatkozások feldolgozása szempontjából, mert a vendégszövegekben szereplő belső hivatkozásokat a befoglaló jogszabály szempontjából külső hivatkozásoknak kell tekinteni, míg a beágyazott jogszabály, a vendégszöveg felől tekintve ezeket belső hivatkozásoknak lehetne minősíteni.

Nem túl gyakran, de előfordul olyan *általános hivatkozás*, amikor nem konkrét, egyedi (partikuláris) jogszabályra utal a jogszabály szövege, hanem valamilyen általános, nem egyedi hivatkozást tesz: *A szexuális szolgáltatás fogalmát és reklámozásának további korlátozásait külön törvény állapítja meg.* A 'külön törvény' típusú utalás annyit jelent, hogy a szóban forgó fogalom (itt a szexuális szolgáltatás fogalma) a jelen törvényben nincs az értelmező rendelkezések között, de a jogrendszer által ismert, kezelt fogalomról van szó. Ezek a hivatkozások formai jellemzőiket tekintve hasonlíthatnak az „igazi” hivatkozásokhoz, és csak azzal lehet elkülöníteni őket, hogy a típus-hivatkozások nem tartalmazzák minden szükséges partikuláris adatot, ami egy konkrét, egyedi jogszabály azonosításához szükséges (tipikus módon hiányzik az időre és a jogszabály konkrét jelére, „sorszámára” vonatkozó információ).

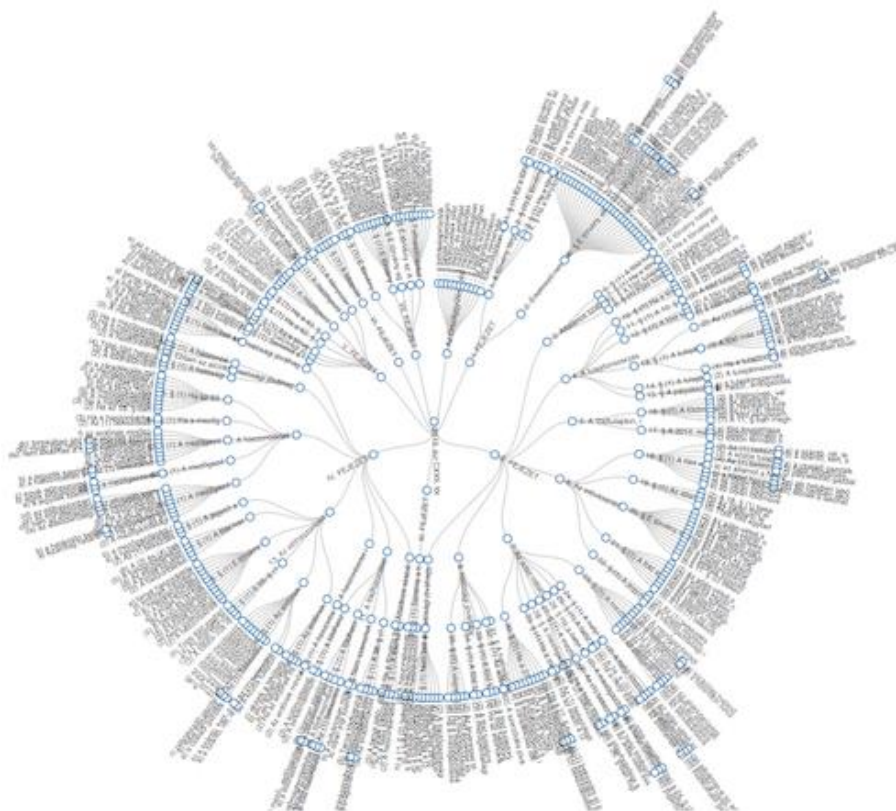
Az előző problémakörhöz hasonlóak azok a jogszabályok, amelyek egyértelműen az elemzett jogszabály megalkotásának időpontját követően létrehozott, létrehozandó jogszabályokra tesznek utalást. Az ilyen hivatkozások ugyanúgy nem teljesek, mint a fentebb bemutatott „általános jogszabályok”, érdemes mégis külön említeni őket, és külön foglalkozni velük, mert ezek a *biankó jogszabályok* idővel jó eséllyel megszületnek, bekerülnek a teljes jogkorpuszba, amikortól fogva már ugyanolyan hivatkozható, partikuláris jogszabályként funkcionálnak, mint a többiek. Tipikus példaként említhetjük a törvények végrehajtási jogszabályait. A biankó jogszabályokat éppen azért érdemes külön kezelni, mert ha az elemzést időben később végezzük el, azután, hogy a biankójogszabály már hatályosították, akkor a biankójogszabály a többi jogszabállyal teljesen „egyenrangú” elemként szerepel a korpuszban, miközben a jelenben még eltérő a státusuk.

Goody-listák és hivatkozások

A jogszabályszövegek jelentős hányada jelölt listákból áll, amelyek elliptikus mondatokként is jellemezhetőek. Ezeket korábban Goody-listáknak neveztük el, és a sajátosságaikat több szempontból is elemeztük [5, 6, 7, 11]. A hivatkozások feltárásban a Goody-listák akkor jelenthetnek gondot, amikor a hivatkozások „átfolynak” a Goody-listák fejtételei és listatételei között. Mivel a Goody-listák elliptikus mondatainak kiegészítését, teljes mondatokká egészítését korábban már megoldottuk, így itt ezt a technikát alkalmazva a probléma könnyen eliminálható. A hivatkozásfeltárás feladatát azon a feldolgozott szövegen érdemes megkezdeni, amely már tartalmazza a Goody-listák hiányos mondatainak kiegészítését.

Vizuális ábrázolás

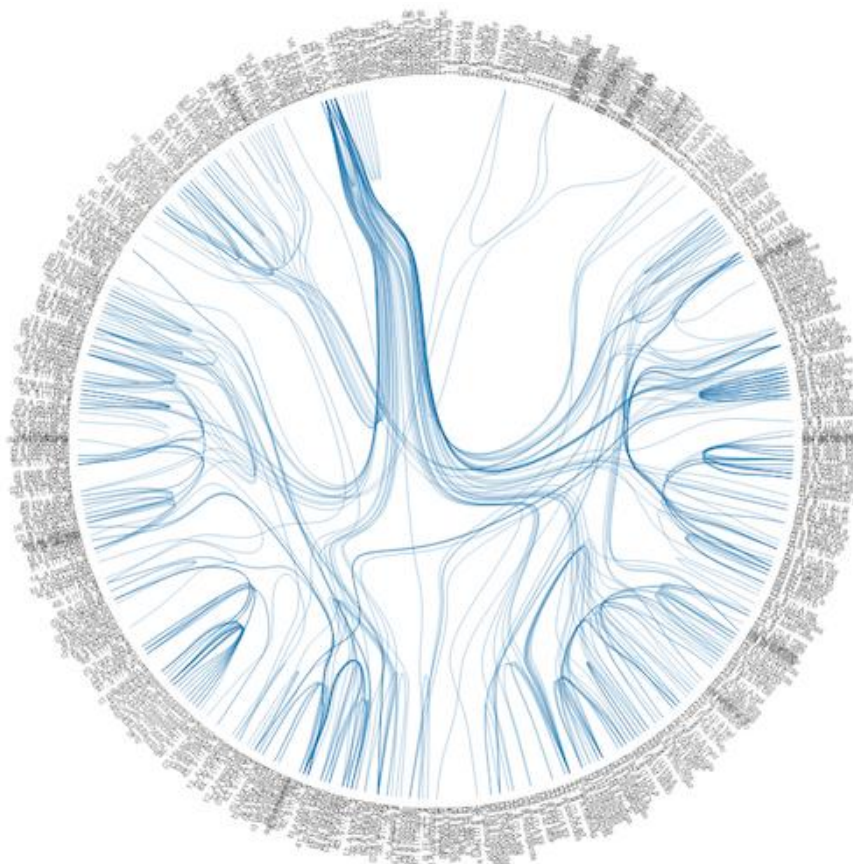
A jogszabályok szövege hierarchikus struktúrájú, ám a hierarchikus elrendezés megszokott szabályaira (pl. a hierarchikus szintek egyenletes eloszlására) a jogszabályok szövegének kialakításakor nem lehetnek figyelemmel. A jogalkotó részletezni/tételezni ott fog, ahol erre tartalmilag (szemantikailag/szintaktikailag) szükség van. Így előfordul, hogy nagyon egyenlőtlen eloszlású hierarchiák jönnek létre, amikor a jogszabályok egyik részében nagyon gazdag alstruktúra jelenik meg, míg máshol kevés szöveg, lapos struktúra a jellemző. A „szép” hierarchikus megjelenítést (és az egyértelmű hivatkozást) nehezíti a – már bemutatott – kettős tipizálás kényszere is (a bekezdések és azok első szakaszai esetében).



1. ábra: a 2013. évi CXXII. törvény szerkezete

A jogszabályok struktúrájának elemzéséhez, a struktúrák tipizálásához kerestünk adatvizualizációs technikákat [13]. A kiválasztott javascript-könyvtárakra [9] támaszkodva elkészítettük a 2013. évi CXXII. törvény [1] belső szerkezetét mutató ábrát (1. ábra). Munkánk következő szakaszában szeretnénk mérhetővé tenni a belső szerkezet jellegzeteségeit, hogy ezek alapján a különböző struktúrátípusok jól elkülöníthetőek legyenek.

Még érdekesebb az az ábra, amely a jogszabályok belső hivatkozásait mutatja (2. ábra). Első ránézésre is kitűnik, hogy a jogszabályon belül sok egymáshoz közeli helyekre mutató, *szomszédos hivatkozás* van (általában a szakaszokon belül vannak így összekötve a bekezdések, pontok, alpontok egymással). Ez a jellegzetesség vélhetőleg minden jogszabály esetében megfigyelhető lesz. Már nem annyira kiterjedten, de még mindig elég sok jogszabályt jellemezhet az az itt megfigyelhető másik vonás, hogy a jogszabály zárórendelkezései között sok visszamutató, belső linket találhatunk. A magyarázat egyszerű: gyakran előfordul, hogy az egyes részek különböző időpontban lépnek hatályba, mely időpontok belső hivatkozással történő felsorolását találjuk meg a jogszabályszöveg lezárásaként. Számos jogszabály belső hivatkozásainak hasonló módon történő ábrázolása összevethetővé teszi a belső linkek strukturáját, mely – munkahipotézisünk szerint – rámutathat bizonyos (pl. jogágankénti vagy jogszabálytípusonkénti) jellegzetességekre, eltérésekre. A belső linkek szerkezetének teljes tipizálásához, a típusok értelmezéséhez még nem áll elég tudás a rendelkezésünkre. Első menetben arra van szükség, hogy sok jogszabály vizualizációját elemezve a tipizáláshoz szükséges szempontokat megtaláljuk, hogy aztán mérhetővé tegyük a típusalkotáshoz szükséges dimenziókat.



2. ábra: a 2013. évi CXXII. törvény belső linkjei

Köszönetnyilvánítás

A tanulmány létrejöttében segített a K112172 sz. OTKA kutatás támogatása. Köszönjük Király Péternek az adatvizualizációhoz nyújtott segítségét.

Hivatkozások

1. A mező- és erdőgazdasági földek forgalmáról szóló 2013. évi CXXII. törvény
2. A jogszabály szerkesztéséről szóló 61/2009. (XII. 14.) IRM rendelet
3. A nemzeti vagyonról szóló 2011. évi CXCVI. törvény
4. Friedl, J.E.F.: *Mastering Regular Expressions*, O'Reilly Media (2006)
5. Hamp, G., Syi, Markovich, R.: Elliptikus listák jogszabálysövegekben. In: Tanács A., Varga V., Vincze V. (szerk.) XI. Magyar Számítógépes Nyelvészeti Konferencia. Szeged (2015) 273–281.
6. Markovich R., Hamp, G., Syi.: Jogszabálysövegek gépi elemzésének tanulságai. *Jogelméleti Szemle*, 2. (2015) 64–73.
7. Markovich, R., Syi, Hamp, G.: Elliptical lists in legislative texts. In: ICAIL '15 Proceedings of the 15th Int. Conf. on Artificial Intelligence and Law (2015) 192–195.
8. Martínez, M., de la Fuente, P., Vicente, D-J.: Dealing with the automatic extraction of references from legislative digital libraries. *Veille Strategique, Scientifique et Technologique 2004 (VSST 2004)* 281–288
9. mbostock's blocks, <http://bl.ocks.org/mbostock>
10. Sartor, G.. Legislative Information and the Web. In Sartor, G., Palmirani, M., Francesconi, E., Biasiotti, M.A. (eds.) *Legislative XML for the Semantic Web*, Springer (2011) 11–20.
11. Syi, Hamp, G., Markovich, R.: Goody-listák a jogszabálysövegekben. *Három tételben. JEL-KÉP 3.* (2015) 13–24.
12. Palmirani, M., Brighi, R., Massini, M.: Automated extraction of normative references in legal texts. In: ICAIL '03, Proceedings of the 9th Int. Conf. on Artificial Intelligence and Law (2003) 105–106.
13. Tufte, Edward R.: *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, CT: Graphics Press (1997)