

Adó- és jövedéki jogi wordnet (TaXWN)

Almási Attila¹, Vincze Veronika¹, Sulyok Márton², Csirik János³

¹ Szegedi Tudományegyetem, Informatikai Tanszékcsoport
Szeged, Árpád tér 2.
vizipal@gmail.com, vinczev@inf.u-szeged.hu

² Szegedi Tudományegyetem, Alkotmányjogi Tanszék
Szeged, Tisza Lajos krt. 54.
msulyok@juris.u-szeged.hu

³ MTA-SZTE, Mesterséges Intelligencia Tanszéki Kutatócsoport
Szeged, Tisza Lajos krt. 103. III. lépcsőház
csirik@inf.u-szeged.hu

Kivonat: A magyar jogi fogalomháló eddig elkészült része (TaXWN) adó és jövedéki tárgykörből tartalmaz mintegy 650 synsetet. Első lépésben a wordnet építéséhez szükséges számítógépes előfeldolgozási munkára került sor. Ezt követte a TaXWN mint adatbázis létrehozása, majd létrehoztuk a synsetjelöltek definícióit és hozzárendeltük a jogi szempontból nélkülözhetetlen megjegyzéseket. Ezután következett a fogalmi háló felépítése. Alapszintű kapcsolódás jött létre a LOIS nemzetközi jogi wordnettel. A rendszer továbbfejleszhető egyrészt a HuWN irányába, másrészt pedig egyéb jogterületek fogalmaival is bővíthető.

1 Bevezetés

A modern információs társadalomban kulcsfontosságú szerepet tölt be az információ gyors megszerzése és hatékony kezelése. A felhasználók széles körének szerteágazó igényei mellett a legkülönbözőbb területeken dolgozó szakemberek számára is elengedhetlenné vált, hogy megfelelő gyorsasággal tudjanak tájékozódni az adott szakirodalomban, az adott probléma megoldásához hatékonyan tudják előkeresni a kapcsolódó háttérismereteket, és egyeztetni tudják egymással a különböző – esetlegesen akár többnyelvű – forrásokból származó információkat. A hatékonyság jellemzői közül a visszakeresett információ terjedelmét, minőségét, relevanciáját, a megszerzéshez szükséges időt és költségeket emeljük ki.

A jogi szövegek számítógéppel támogatott feldolgozásához és kutatásához szükséges a jogi szakszövegek digitalizálása, az elektronikus formában való nyilvánossá tétele és az elektronikus dokumentumkezelés biztosítása. Ez utóbbi előfeltétele egy, az adott szakterület fogalomkészletét lefedő ontológia [8].

A magyar jogi wordnet kezdeményezés a magyar jogrendszer nemzetközi (közelebbről európai uniós) jogrendbe történő integrációjában tölthet be fontos szerepet, hiszen egy olyan jogi tudásbázist alakít ki, amely az uniós jogközelítés, jogharmonizáció lexikális háttérét adja, ezáltal megkönnyíti a szükséges változtatások elvégzését. A cél tehát az EuroWordNet [2] eredményeire és formalizmusára építő, szemantikai-

lag strukturált, jogi vonatkozású fogalomtár létrehozása a magyar nyelvre, melyet a jog, az informatika és nyelvészet tudományágai együttműködve alkotnak meg. Ennek első lépéseként az adó- és jövedéki jogi wordnet (TaXWN) készült el.

2 A magyar jogi wordnettel kapcsolatos informatikai feladatok

Első lépésben a wordnet építéséhez szükséges számítógépes előfeldolgozási munkálatokra került sor. A rendelkezésre álló jogszabályok, illetve egyéb szakirodalmi anyagok feldolgozásával terminológiajelölt listák kerültek előállításra, majd a továbbiakban ezek felhasználásával, lényegében emberi erővel történt meg a tényleges adatbázis-fejlesztés.

Az informatikai feladatok a wordnet kézi építése során döntően különböző formai (szólisták konverziója az adatbázis formátumára, automatikusan kitölthető adatbázis mezők felvitele, stb.), illetve validációs (az adatbázis formalizmusa által támasztott követelményeknek meg nem felelő fogalmak kiszűrése, szintaktikailag hibás fogalmak szűrése, stb.) munkákra korlátozódtak. A felmerült feladatok elvégzéséhez, valamint az adatbázis automatikus minőségbiztosítási mechanizmusának kifejlesztéséhez elkészült egy, az adatbázis formátumát felolvasni képes osztálykönyvtár, melyben az egyes – a nyelvész kollégák által megfogalmazott, illetve strukturális (XML-validáció) – szűrések gyorsan és egyszerűen megvalósíthatók voltak.

3 A magyar jogi wordnettel kapcsolatos jogi feladatok

A projektnek az SZTE Állam- és Jogtudományi Kar Alkotmányjogi Tanszék részéről koordinált részének célkitűzései a következők voltak:

- a) az adózás rendjéről szóló, 2003. évi XCII, illetve az adózás és jövedék témakörökhöz kapcsolódó egyéb jogszabályok szakterminológiájának kivonatolása;
- b) egy – elsősorban eljárási, jogalkalmazói felhasználási célhoz igazodó – értelmező szótár készítése a fenti kivonat alapján.

A kivonatolandó fogalmi kört először úgy próbáltuk meghatározni, hogy vettük az adott témához kapcsolódó irodalmi terminológiát (a továbbiakban: lit-források), illetve a kapcsolódó jogszabályok terminológiáját (a továbbiakban: lex-források). Miután nyilvánvalóvá vált, hogy a két forrásanyag között olyan eltérések lehetnek, amelyek a fentebb leírt alapcélokat ellehetetlenítik, így a későbbiekben el kellett tekintenünk a lit-források használatától¹. Ennek oka az volt, hogy az egyes lit-források gyakorta nem azonos definíciót használtak az egyes lex-forrásokban fellelhető fogalmakra, ami egyrészt a szerzői szabadság és szubjektivitás, másrészt pedig az oktatási célra fel-

¹ A későbbiekben a lex-forrásokból nyert fogalmak értelmetlensége és esetleges hiányossága esetén pontosításokra, egyértelműsítésre használtuk fel a lit-források anyagait számos esetben, a jogalkalmazó munkájának megkönnyítése céljából.

használandó szellemi alkotásoktól elvárt közérthetőség követelményéből adódik, s ez gyakran a szakkifejezések egyszerűsítését, átfogalmazását igényli. A jogalkalmazó azonban elsősorban a jogforrásokban, jogszabályokban fellelhető fogalmakra támaszkodik munkája során, tehát ezek elsőbbsége egy jogi wordnet kialakításában indokolt.

Mégis felhasznált irodalmi forrás vagy saját elképzelések alapján átalakított definíciók esetében kiegészítő információként (*Megjegyzés / Note*) jelöltük a hivatkozott forrást, illetve „egyéni” címkével látjuk el az adott meghatározást (vö. 4.1 és 4.2).

A jogi wordnet alapjául szolgáló szakterminológia-kivonat XLS formátumú, ún. *LEXtract* (jogszabályi kivonat) listákba rendezve az alábbi elemeket tartalmazta: a kifejezéseket, azok definícióit és a többletinformációt tartalmazó megjegyzéseket.

1. táblázat: TaXWN LEXtract

TaXWN_LEXtract		
<i>Kifejezés</i>	<i>Definíció</i>	<i>Megjegyzés</i>
adatgyűjtésre irányuló ellenőrzés	Olyan ellenőrzési eljárás, amelynek célja az adóhatóság nyilvántartásában és az adózó nyilvántartásában, bevallásában szereplő adatok, tények, körülmények valóságtartalmának, illetőleg ezek hitelességének megállapítása.	Art. 119 § (1); Az adatgyűjtésre irányuló ellenőrzés során az adóhatóság a bevallási időszak lezárását megelőzően is adatokat gyűjthet.
adatkérés	A kapcsolattartó közigazgatási szervhez az Európai Közösség tagállami illetékes hatósága által küldött olyan kérelem, amelyben ez utóbbi a tartozás behajtásához szükséges adatok átadását kéri.	Art. 61 § (1)
adóbevallás	Az adózó azonosításához, az adóalap, a mentességek, a kedvezmények, az adó, a költségvetési támogatás alapja és összege megállapításához szükséges adatokat tartalmazó nyilatkozat.	Art. 31. § (1)-(14); Jöt. 48./B§ (2) Az adóalany adóbevallási kötelezettségét elektronikus úton a külön jogszabályban foglalt módon és technikai feltételekkel teljesíti.

4 A magyar jogi wordnettel kapcsolatos nyelvészeti feladatok

A szűrés, listázás és átválogatás után megmaradt fogalomjelölteket synsetekbe rendeztük. Ezt követően került sor a definíciók, megjegyzések, valamint a hierarchia létrehozására. A szakontológia építésével kapcsolatban itt kell megemlítenünk, hogy mivel a jelen jogi wordnet adó- és jövedéki ontológiája túlzottan specifikus, annak synsetjei gyakran egyeleműek, ami az általános wordnetek (például HuWN, PWN) esetében ritkaságnak számít.

4.1 A definíciók létrehozása

A nyelvészet és a jog által támasztott követelmények gyakran ellentmondásba kerültek, ezért ki kellett mondanunk, hogy a rendszer építésénél csak az egyik tudományterület (jelen esetben a jog) igényeinek felelhetünk meg teljesen, de amennyiben lehetséges, a nyelvészet követelményeit is megpróbáljuk figyelembe venni.

A feladat kezdetén tehát rögzítettük, hogy a rendszer egy jogi alapon álló fogalmi háló (wordnet) lesz. Ennek következtében például módosult az a wordnetépítésben megszokott nyelvi szabály is, miszerint egy fogalom definíciójának tartalmaznia kell a fogalom egy hipernimáját (egy általánosabb fogalmat) vagy annak valamely szinonimáját [1]. Ez az esetek nagy részében nem így történt, mivel a definíciók – amelyeket jogász szakértők állítottak össze törvények szövegeire támaszkodva – gyakran csupán fõlsorolások, melyeknek egyes elemei egy nyelvészeti szempontból kielégítő hálóban csak meronimák lehetnek volna.

Így rendszerünk olyan definíciókat tartalmaz, amelyek teljes mértékben kielégítik a jogtudomány támasztotta igényeket, és esetenként a nyelvészeti elvárásoknak is megfelelnek.

4.2 A megjegyzések kialakítása

A magyar wordnet (HuWN) [4, 5] létrehozásakor a *Megjegyzés* a synseten belül egy olyan egység volt, amely a megállapodás szerint rövid kiegészítő megjegyzések felvételét tette lehetővé. Itt elsõsorban a Princeton WordNetben (PWN) [1] és a HuWN-ben eltérõ szõfajúként megjelenõ synseteket jelöltük, valamint ide kerültek a gazdasági szakontológia synsetjeit jelölõ „szak” megjegyzések. Ezekon kívül, a javítási fázist megelőzően lehetőség volt arra, hogy a synset létrehozását végző nyelvész saját megjegyzéseit is felvehesse, amely a késõbbiekbõn támpontot nyújtott a javítást végző munkatársaknak.

A jogi szakontológiában azonban a *Megjegyzés* egy teljesen eltérõ funkcióval rendelkezik. Ide kerültek azok az információk, amelyek magába a definícióba nem fértek be, de a meghatározandó fogalommal kapcsolatban olyan adatokat tartalmaznak, amelyek szerepeltetése nélkül a kívánt jogi tartalom nem lenne teljes. Ezenkívül itt találhatóak azok a kiegészítések is, amelyek arról nyújtanak információt, hogy az adott fogalom mely törvényben lett szabályozva, hogy milyen számszerû adatok lehetnek lényegesek a jövendõ felhasználó számára a fogalommal kapcsolatban (pl. alkoholfok, importálható árucikk mennyisége stb.).

4.3 A hierarchia létrehozása

A TaXWN megalkotásának lényegi mozzanata volt a fogalmi háló felépítése. Ennek létrehozásakor nem támaszkodhattunk egy már elkészült rendszerre, mint pl. a HuWN esetében, s ez egyszerre jelentett könnyebbséget és nehézséget is. Könnyebbséget azért jelentett, mert a hierarchia létrehozásánál nem kellett azzal törődnünk, hogy az egy másik rendszerrel összevethető, esetleg összekapcsolható legyen. Nehezebb azért volt, mert ebben az esetben saját kútfőre támaszkodva kellett egy használható, értelmes hierarchiát fölállítanunk.

Újabb csomópontok

A hierarchia kialakítása során a *bottom-up* módszert követtük, mert a törvényi forrásokból származó anyag igen specifikus kifejezések, s ezáltal általában csak alapsynsetek létrehozását tette lehetővé. Ezzel egyébként a munka egyszerűbbé is vált, mivel a jogi domént elhagyva a hipernimákat legtöbbször már a HuWN synsetjeinek és hierarchiájára támaszkodva tudtuk kiválasztani.

Az ügynevezett *unique beginner*

Rendszerünkben kilenc *unique beginner* synset található, amelyek a hierarchia legáltalánosabb synsetjei. Ezeknek a synseteknek a megtalálása általában magától értetődő volt, máskor viszont hosszadalmas utánajárást igényelt. Ennek oka először is a hierarchiaépítés megegyezés szerinti első szabályában keresendő, miszerint a jogi wordnet egy, a jogi szakszókincs által körülhatárolt háló lesz és a rendszer kialakítása során eltekintünk attól, hogy minden esetben a specifikusabb nyelvészeti szempontokat vegyük figyelembe. Így fordulhatott többször elő, hogy egy, az alapsynset szintjén még tárgyként azonosított elem magasabb szinten egy nem tárgyként azonosítható hipernima alá, a legfőbb szinten pedig akár az *elvont fogalom* vagy *állapot* alá került bekötésre. Azonban jogi wordnetünkben nem lehetséges az összes ilyen, nyelvészeti szempontból úgymond lehetetlen állapot megszüntetése. A jog nyelvzetéből és a fentebb említett megállapodásból eredően ezek a látszólagos „következetlenségek” meg kell, hogy maradjanak.

5 Kapcsolódás a LOIS-hoz

A LOIS Project, vagyis a *Lexical Ontologies for Legal Information Sharing* rendszere hatékony, európai szintű, információs-kommunikációs, fejlesztési együttműködést céloz meg. Ez a program a EuroWN-en keresztül kapcsol össze 6 különböző tagállami wordnetet (cseh, angol, német, holland, portugál és olasz) [6, 7].

Az SZTE Informatikai Tanszékcsoport és a LOIS konzorcium vezető intézménye, az *Institute of Legal Information Theory and Techniques* között létrejött megállapodás alapján vállaltuk, hogy megvizsgáljuk a jogi wordnet LOIS-hoz való kapcsolásának lehetőségét, és a kutatás eredményeit megosztjuk a LOIS projekt felelőseivel.

A LOIS megközelítőleg 7000 kifejezést tartalmazó angol nyelvű általános jogi terminológiájának XML-fájljából kivontuk a LEMMA=”kifejezés” sorokat (ebben szerepeltek a magyar terminológiával összehasonlítható szavak), majd azokat össze-

vetettük a magyar jogi wordnet mintegy 650 kifejezésből álló adó- és jövedéki terminológiájával.

Maga az összekapcsolás a következő módon történt: az egyező synsetek, fogalmak LOIS-ban található azonosítóját fölvevük – a TaXWN építésénél használt VisDic szerkesztő [3] segítségével – a megfelelő synset *Megjegyzés* ablakába, mégpedig a következő formában: LOIS ID="xxx".

A kapcsolódási pontok megtalálása nem volt egyszerű feladat, mert az a magyar jogi terminológia, amely a céltartomány (adó, jövedék) kiegészítéseként kellett bekerülnön a TaXWN rendszerébe – a fogalmi pontosság és a megfelelő synset kapcsolatok kialakítása végett – nem minden esetben volt pontosan megfeleltethető az uniós (ez esetben angol nyelvű) terminológiának. Például:

<p>A LOIS-ban szereplő kifejezések: <i>company, business, undertaking, enterprise, firm, corporation, concern, business corporation</i></p> <p>A TaXWNkifejezései: <i>társaság, cég, vállalkozás, vállalat, vállalkozási tevékenység</i></p>
--

1. ábra. A LOIS és a TaXWN kifejezéseinek megfeleltetése.

A nyolc angol kifejezéseknek öt magyar felel meg valamilyen mértékben, de az egy az egyhez való megfeleltetés jogi szempontból nem lehetséges.

A kapcsolódás gyakran azért sem lehetséges, mert a LOIS definícióiban sok helyütt úgy utal az egyes kifejezésekre, mint pl. „ezen Egyezmény értelmében” vagy „a 25. cikkben írtakkal összhangban”. Ezek a kontextusok a mi terminológiánkra vonatkoztatva elvesztik értelmüket, mivel anyagunk nem tartalmazza a közösségi terminológia által használt kifejezéseket, illetve forrásanyagokat.

Nyelvész és jogász munkatársaink tehát megvizsgálták az összes kifejezés angol és magyar nyelvű definícióját, hogy a lehető legmegfelelőbb kapcsolatokat alakíthassák ki.

Nehézséget okozott még, hogy a LOIS nem minden esetben ad meg definíciót és/vagy példát az adott synsethez, ezért ezeknek a synseteknek a minden kétséget kizáró azonosítására nem volt lehetőségünk, s így nem is vettük föl őket a TaXWN-be.

Előfordult, hogy a TaXWN-ben szereplő synset egyszerre több LOIS synsethez is kapcsolható volt. Ez arra hívja föl a figyelmet, hogy ilyen specifikus, magyar nyelvű szakszóanyagot nem lehet minden esetben egyértelműen megfeleltetni például egy általánosabb, ebben az esetben angol nyelvű jogi kifejezésnek.

Az összekapcsolás nehézségeinek bemutatására áll itt a *minor/juvenile* és a *kiskorú/fiatalkorú* kifejezések példája.

A LOIS-ban:

```
<WORD_MEANING          ID="1429"          PART_OF_SPEECH="N"
STATUS="FINISHED">
<SOURCEBASE>LEXDB</SOURCEBASE>
<NOTE/>
```

```

<GLOSS>a person who has not reached full legal
age</GLOSS>
<CONCEPTS/>
<VARIANTS>
<LITERAL LEMMA="minor" SENSE="1">
<EXAMPLES>not of legal age; &quot;minor
children&quot;</EXAMPLES>
</LITERAL>
<LITERAL LEMMA="minor" SENSE="1">
<EXAMPLES>a person who has not reached full legal age;
a child or juvenile</EXAMPLES>
</LITERAL>
<LITERAL LEMMA="juvenile" SENSE="1">
<EXAMPLES>a person who has not reached the age
(usually 18) at which one should be treated as an adult
by the criminal justice system</EXAMPLES>
</LITERAL>
</VARIANTS>
<INTERNAL_LINKS>
<RELATION ID="3089" TYPE="involved_patient">
<TARGET_WM ID="379" PART_OF_SPEECH="N"/>
</RELATION>
</INTERNAL_LINKS>
<EQ_LINKS>
<RELATION ID="14290000" TYPE="eq_has_hyperonym">
<TARGET_WM ID="2858" PART_OF_SPEECH="N"/>
</RELATION>
</EQ_LINKS>
</WORD_MEANING>

```

A TaXWN-ben:

```

<SYNSET>
<ID>HuWN-1665382021</ID>
<SYNONYM>
<LITERAL>kiskorú
<SENSE>0</SENSE>
</LITERAL>
</SYNONYM>
<DEF>A. 18. életévét be nem töltött személy, kivéve, ha házasságot kö-
tött.</DEF>
<SNOTE></SNOTE>
<SNOTE>Art.35.§ (2) i) A szülői felügyelet alatt álló kiskorú adótarozásáért a
szülői felügyeletet gyakorló szülő az általa kezelt va</SNOTE>
<SNOTE>Art.5.§ (2) b</SNOTE>
<SNOTE>LOIS ID="1429"; a magyar jogrendben kis- és fiatalokú megkülön-
böztes l</SNOTE>
<SNOTE>jog</SNOTE>
<POS>n</POS>

```

<ILR>HuWN-148541600
 <TYPE>hypernym</TYPE>
 </ILR>
 <STAMP>almasi 2008/12/02</STAMP>
 </SYNSET>

<SYNSET>
 <ID>HuWN-911671085</ID>
 <SYNONYM>
 <LITERAL>**fiatalkorú**
 <SENSE>0</SENSE>
 </LITERAL>
 </SYNONYM>
 <DEF>Fiatalkorú az, aki a bűncselekmény elkövetésekor tizennegyedik életévét betöltötte, de a tizennyolcadikat még nem.</DEF>
 <SNOTE>1978. évi IV. tv. Btk. 107.§. (</SNOTE>
 <SNOTE>LOIS ID="1429"; a magyar jogrendben kis- és fiatalkorú megkülönböztetés l</SNOTE>
 <SNOTE>jog</SNOTE>
 <POS>n</POS>
 <ILR>HuWN-148541600
 <TYPE>hypernym</TYPE>
 </ILR>
 <STAMP>almasi 2008/12/02</STAMP>
 </SYNSET>

A TaXWN-ben a *kiskorú*: **A 18. életévét be nem töltött személy, kivéve, ha házasságot kötött**; míg a *fiatalkorú*: **Az, aki a bűncselekmény elkövetésekor tizennegyedik életévét betöltötte, de a tizennyolcadikat még nem**. Tehát, amíg a LOIS-ban a *minor* (kiskorú) kifejezésbe jogilag beletartozik a *juvenile* (fiatalkorú) kifejezés is, addig a TaXWN azokat külön kezeli, saját definíciókkal. Más kérdés, hogy a LOIS-ban a *juvenile* kifejezésnek van saját definíciója, noha a *minor*-é eleve magában foglalja a *juvenile*-t is. A magyar büntetőjogi terminológia viszont nem a kiskorúak részeként kezeli a fiatalkorúakat, hanem jogilag külön kategóriába sorolja őket. Ezt a két fogalmat végül mi is külön vettük föl és a *Megjegyzés*-ben jelöltük, hogy jogi értelemben a magyar terminológia hogyan különbözteti meg a *kiskorú* és *fiatalkorú* kifejezéseket.

6 Statisztika

2. táblázat: A TaXWN synsetjeinek megoszlása.

	LOIS-hoz kapcsolható	LOIS-hoz nem kapcsolható	TaXWN
általános	81	116	197
jövedéki	113	337	450
összesen	194	453	647

Amint látható, a TaXWN 647 synsetje közül biztonsággal 194-et lehetett a LOIS nemzetközi jogi ontológia synsetjeihez kapcsolni. Ebből 113 szorosan kapcsolódik a jövedéki terminológiához, 81 pedig általánosabb tartalmú kifejezés.

A TaXWN-en belül az adó és jövedéki területhez szorosan kapcsolódó, törvényi anyagokkal megtámogatott synsetek száma 450, az általánosabb kifejezéseket tartalmazó synseteké pedig 197. A 647 synset egy kilenc fából álló fogalomhálót képez a rendszerben meglévő *unique beginner* synsetek (*állapot, cselekmény, cselekvés, együttes/összesség, elvont fogalom, entitás, hely, jelenség, tulajdonság*) alapján. Azonban gyakran előfordul, hogy már a jogi tartalmú synset hipernimája elhagyja a szűken értelmezett jövedéki domént és egy általános fogalmi hálóba illő synsetnek tekinthető.

7 A rendszer bővítésének, frissítésének lehetőségei

Először is a magyar jogi wordnet eddig elkészült részét tovább bővíthetjük más törvényi anyagok (pl. büntetőjog, polgári jog, stb.) kifejezéseivel és egy összetettebb hierarchiába szervezhetjük őket.

Másodsor, a magyar jogi wordnet és HuWN összekapcsolása kétféleképpen is elképzelhető:

1. A HuWN-ben már meglévő jogi tartalmú synsetek összehangolása a magyar jogi wordnettel

Itt a következő probléma adódik. A HuWN-be korábban fölvetett, jogi tartalmú synsetek rendszerint nem ütik meg azt a mércét, amivel egy jogi fogalomtárnak rendelkeznie kell. Ennek egyik oka az, hogy a HuWN elkészítésében nem vett részt jogász szakértő, másrészt azt is érdemes újra megjegyezni, hogy a HuWN synsetjeinek döntő többsége a PWN-ből került átvételre, lett lefordítva. A fordítási nehézségeken túl gondot okoz még az is, hogy a PWN-be az angolszász jogrend szakkifejezései kerültek be. Ezeknek nem mindig van megfelelője a magyar (és/vagy uniós) jogrendben, vagy pedig nem egészen azonos a jogi tartalmuk, ezért ezen problémák kiküszöbölésére jogász szakértő segítségét kell igénybe venni.

2. A magyar jogi wordnet általános (nem jövedéki jogi) synsetjeinek összekapcsolása a HuWN-nel

Mint fentebb említettük, a jogi ontológia egy igen szűk tartományban mozog és azt elhagyva, magasabb szinteken már átválthat az általános magyar nyelvi ontológiába, így a magyar jogi wordnet és a HuWN összekapcsolása könnyen megvalósítható.

Harmadszor, a LOIS teljes anyagát át lehetne ültetni magyar nyelvre, ezáltal Magyarország is teljes mértékben kapcsolódhat a nemzetközi jogi adatbázishoz.

Negyedszer, a rendszer frissítése bizonyos időközönként szükségesnek bizonyul, ami a jogalkalmazó munkája szempontjából mindenképpen elengedhetetlen. Az egyes jogszabályok esetleges megváltozása így könnyen nyomon követhetővé válik a hazánkban ismert nagyobb elektronikus jogszabálygyűjteményeknek, illetve online adatbázisoknak a rendszerbe történő automatikus integrálása által.

8 A rendszer felhasználhatóságának lehetőségei

A magyar jogi wordnet az elektronikus dokumentumkezelés ideális kiegészítő programja lehet, amely, többek között, például a LOIS Project adatbázisain keresztül megteremtheti a szükséges többnyelvű tudásbázist ahhoz, hogy az adott tagállamban eljáró szerv az általa nem ismert nyelven kitöltött dokumentumokat is könnyen értelmezhesse, kezelhesse, és ezzel az ügyintézési időt lerövidítse. A jogi korpusz egyéb felhasználási területei lehetnek még például az emberi ellenőrzés mellett futó, de alapvetően automatikus szakfordító programként történő alkalmazás, vagy éppen egy webalapú, bárki számára hozzáférhető, jogi szakszótári funkció is.

9 Összefoglalás

A TaXWN létrehozásával egy általános magyar jogi wordnet első lépcsője, az adó- és jövedéki alapterminológiát tartalmazó rész valósult meg. A fogalmi háló létrehozása után következő munkálatok magukban foglalták a LOIS jogi wordnethez való csatlakozás lehetőségének megteremtését, ami először is a TaXWN és a LOIS angol nyelvű változatában szereplő fogalmak, synsetek összevetését jelentette. Ezt követte azután a LOIS-ban egyezést mutató synsetek azonosítóinak fölvétele – a nyelvközi index (ILI) segítségével – a TaXWN-be. Jelen állapotban az egyirányú indexelés miatt csak a magyar jogi wordnet felől érhetők el a LOIS synsetjei. Az elkészült rendszer nagyságrendjében és minőségében elérte a kezdeti elvárásokat és jó elméleti és gyakorlati alapot nyújthat egy következő lépésben létrehozandó egyéb jogterületet lefedő fogalmi hálónak. A HuWN jogi synsetjeinek javítása és a LOIS jogi wordnethez történő, magasabb szintű, átfogóbb csatlakozás pedig későbbi projektek témája lehet.

Köszönetnyilvánítás

A kutatást – részben – a TUDORKA és a MASZEKER projekt (Jedlik Ányos programok) keretében az NKTH támogatta.

Hivatkozások

1. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: Introduction to WordNet: an On-line Lexical Database. *International Journal of Lexicography*, Vol. 3, No. 4 (1990) 235–244
2. Alonge, A., Bloksma, L., Calzolari, N., Castellon, I., Marti, T., Peters, W., Vossen P.: The Linguistic Design of the EuroWordNet Database, *Computers and the Humanities. Special Issue on EuroWordNet*, Vol. 32, No. 2–3 (1998) 91–115
3. Horák, A., Smrž, P.: New Features of Wordnet Editor VisDic. *Romanian Journal of Information Science and Technology. Special Issue on BalkaNet*, Vol. 7, No. 1–2 (2004) 201–213
4. Alexin, Z., Csirik, J., Kocsor, A., Miháltz, M., Szarvas, Gy.: Construction of the Hungarian EuroWordNet Ontology and its Application to Information Extraction. Project report. In: *Proceedings of the Third International WordNet Conference (GWC2006)*, January 22–26, South Jeju Island, Korea (2006) 291–292
5. Miháltz, M., Hatvani, Cs., Kuti, J., Szarvas, Gy., Csirik, J., Prószéky, G., Váradi, T.: Methods and Results of the Hungarian WordNet Project. In: Tanács, A., Csendes, D., Vincze, V., Fellbaum, C., Vossen, P. (eds.): *Proceedings of the Fourth Global WordNet Conference. GWC 2008*, University of Szeged, Department of Informatics (2008) 311–320
6. Sagri, T., Tiscornia, D.: Semantic Lexicons for Accessing Legal Information. In: Traunmüller, R. (ed.): *Electronic Government, Third International Conference, EGOV 2004 Proceedings (Zaragoza, Spain, 2004 30 August - 4 September)* (2004) 72–81
7. Peters, W.: The LOIS Project. In: Sojka, P., Choi, K.-S., Fellbaum, C., Vossen, P. (eds.): *GWC 2006, Proceedings* (2006) 331–332
8. Sulyok M., Gyenge B.: Jog és nyelv kapcsolata egy nem mindennapi vállalkozásban. *Közjogi Szemle*, 2. évfolyam, 2009. szeptember, (2009) 49–60