



**Budapesti Műszaki és Gazdaságtudományi Egyetem**

Villamosmérnöki és Informatikai Kar

Méréstechnika és Információs Rendszerek Tanszék

Boros Dávid

# **JOGI SZAKÉRTŐI RENDSZEREK FEJLESZTÉSE**

önálló labor

KONZULENS

Dr. Strausz György

Förhécz András

BUDAPEST, 2012

## Tartalom

1.	Feladatkiírás .....	1
1.1.	Féléves projekt .....	1
2.	A „linked data” elképzelés .....	2
2.1.	Szemantikus web.....	2
2.2.	XML.....	2
2.3.	RDF.....	3
2.4.	SKOS .....	3
2.5.	OWL .....	4
3.	Jogi szövegek .....	4
3.1.	Metalex CEN.....	5
3.2.	TVSZ.....	5
3.3.	EuroVoc .....	6
4.	Nyelvi elemzés.....	7
4.1.	UIMA.....	7
4.1.1.	Sentence Splitter.....	8
4.1.2.	Lemmatizer.....	9
4.1.3.	POS tagger.....	9
5.	Eredmény .....	10
5.1.	Folyamat .....	10
5.2.	Link Model.....	11
5.3.	Emerald WebVisualizer .....	12
6.	Továbbfejlesztési irányok .....	14
6.1.	Pontosság javítás .....	14
6.2.	Elemzés bővítése.....	14
6.3.	Szemantikai tartalom.....	14
7.	Irodalomjegyzék.....	15

## 1. Feladatkiírás

A jogi szakértői rendszerek fejlesztése egy alapvető társadalmi igényt, a környezetünket meghatározó szabályozási környezet megismerését, áttekinthetőségét, a jobb eligazodás lehetőségét kívánja támogatni. A megoldások általában olyan döntéstámogató eszközök, amely a tárgyterület tudásalapú modellje alapján képes a felhasználóval kommunikálni, döntéseket hozni, illetve javaslatairól jogszabály alapú magyarázatot adni.

A legfontosabb fejlesztési irányok a területen:

- szabályozás vagy jogszabály és a tárgyterület tudásalapú modellje alapján tanácsadás vagy döntéstámogatás biztosítása
- szabályozás vagy jogi dokumentumok elemzése, tartalmi kereshetőség biztosítása (XML, nyelvtechnológiai eszközök)
- jogi munkafolyamatok támogatása

Egy tárgyterületről rendelkezésünkre álló információ és tudás rendezésének, strukturálásának hatékony megközelítési módja az ontológiai megközelítés. A tárgyterület fogalmainak és ezek szemantikai viszonyainak leírásán alapuló tudástechnológia megfelelő alapot ad a tudás modellezésén túl a szakterületi konszenzus kialakításához, a szakértelem újrafelhasználhatóságának biztosításához. A fogalmi alapú megközelítés és a hozzájuk kapcsolódó technológiák újszerű lehetőségeket biztosítanak a tárgyterületi tudás és ezek forrásai (általában forrásdokumentumai) közötti közvetlen kapcsolat leírásához. A jogszabályok elemezhetőségének, kereshetőségének biztosításához erre a célra kidolgozott XML szabványt célszerű alkalmazni, a dokumentumokat XML dokumentumtárakba menteni. Fontos feladata a jelenlegi fejlesztéseknek a dokumentumok feldolgozása és az átalakított dokumentumok és a dokumentumok tárgyát leíró tudásmodellek kapcsolatának létrehozása, keresése, felhasználó szolgáltatásokhoz felhasználása.

### 1.1. Féléves projekt

A félév során egy közös projektben vettem részt, amely keretén belül a TVSZ<sup>1</sup> bizonyos paragrafusaira készült el egy dialógus rendszer a hozzá kapcsolódó ontológiával és dokumentumelemzéssel.

Ezek közül az utóbbi volt az én feladatom és az ennek a megvalósítása során felhasznált eszközöket és folyamatokat fogom bemutatni ebben a dolgozatban.

## 2. A „linked data” elképzelés

Az emberiség temérdek írásos anyaggal rendelkezik, amelyek nagy része bárki számára hozzáférhető (akár a világhálón, digitális formátumban is), ám értelmezése gyakran nehéz még a hozzáértők (nemhogy egy gép) számára is. A különböző dokumentumokban szereplő azonos fogalmak, adatok nincsenek összekapcsolva, a szinonimák nem mindig azonosíthatóak egyértelműen és természetes problémát jelent a dokumentumok nyelvének, nyelvezetének különbözősége is.

Ennek a problémának a megoldására Tim Berners Lee, a WWW atyja<sup>2</sup> javasolta a linked data<sup>3</sup> elképzelést, amely a különböző, strukturált adatokat kapcsolja egymással össze. Négy fő alapelv<sup>4</sup> a következő:

- A dolgok azonosítására használjunk URI-kat (Unified Resource Identifier).
- Használjunk HTTP alapú URI-kat, így hivatkozások és kapcsolatok könnyen feloldhatóak mind ember és gép számára is.
- A hivatkozás feloldáshoz és információk hozzáadásához jól definiált szabványokat és eszközöket használjunk, mint pl. RDF, OWL (ezekről később részletesebben is írok).
- Adjuk meg a linkeket más, kapcsolódó URI-khoz, ezzel növelve a szolgáltatott információ mértékét.

### 2.1. Szemantikus web

Erre a gondolatra épül a szemantikus világháló elmélete<sup>5</sup>. Célja az, hogy az interneten elérhető temérdek információt, és adatok számítógépek által feldolgozhatóvá és értelmezhetővé tegye. Ezt jól formált adatstruktúrával (XML, RDF, SKOS), metainformációk bevitelével, és különböző, az adatokhoz kapcsolódó fogalmi rendszerek hozzáadásával (és az ezen végzett következtetésekkel) törekszik elérni.

### 2.2. XML

Az XML<sup>6</sup> (Extensible Markup Language) a W3C által javasolt általános célú leírónyelv. Több tíz éve használják már, szigorú szintaktikára épül és platform független. Ezen tulajdonságok teszik ideálissá dokumentumok tárolására és szállítására valamint könnyítik a gépi feldolgozását, algoritmikus elemzését is a szövegeknek.

### 2.3. RDF

Az RDF<sup>7</sup> (Resource Description Framework) az XML-re épülő leíró nyelv, a fő célja, hogy az erőforrásokhoz metaadatokat kapcsoljon, a szöveg gép általi megértést segítse ezzel. Az erőforrásokat a már korábban említett HTTP URI-k jelölik.

Az URI általában webes link, amely egyértelműen azonosít egy erőforrást. Állításokat fogalmazhatunk meg a használatukkal, információkat csatolhatunk hozzájuk. A legfontosabb azonban, hogy a segítségükkel a különböző helyeken, más-más dokumentumokban fellelhető, azonos dolgok (és a hozzájuk kapcsolódó állítások) összekapcsolhatóak.

Példa RDF szintaktikára:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:cd="http://www.recshop.fake/cd#">

  <rdf:Description rdf:about="http://www.recshop.fake/cd/Empire Burlesque">
    <cd:artist>Bob Dylan</cd:artist>
    <cd:country>USA</cd:country>
    <cd:company>Columbia</cd:company>
    <cd:price>10.90</cd:price>
    <cd:year>1985</cd:year>
  </rdf:Description>

</rdf:RDF>
```

### 2.4. SKOS

Az SKOS<sup>8</sup> (Simple Knowledge Organization System) az RDF-re épül, főként teauruszok, taxonómiák és egyéb strukturált szótárak leírására szolgál.

Példa SKOS szintaktikára:

```
<skos:Concept rdf:about="http://example.com/Concept/0001">
  <skos:prefLabel>English cuisine</skos:prefLabel>
  <skos:altLabel>English dishes</skos:altLabel>
  <skos:altLabel xml:lang="fr">Cuisine anglaise</skos:altLabel>
  <skos:inScheme rdf:resource="http://example.com/thesaurus"/>
  <skos:broader rdf:resource="http://example.com/Concept/0002"/>
```

```
<skos:related rdf:resource="http://example.com/Concept/0003"/>
</skos:Concept>
```

## 2.5. OWL

Az OWL<sup>9</sup> (Web Ontology Language) leíró logikákon alapuló ontológianyelv. A szemantikus web része, abban az esetekben igazán hasznos, amikor az RDF keretei már nem elegendőek az információ és tudás kifejezésére. Segítségével fogalmakat, azok kapcsolatát és rajtuk értelmezett axiómákat írhatunk le. Ezen fogalmak természetesen összekapcsolhatóak az RDF-ből megismert URI-kkal.

Példa OWL szintaktikára:

```
<owlx:Class owlx:name="Wine" owlx:complete="false">
  <owlx:Annotation>
    <owlx:Label xml:lang="en">wine</owlx:Label>
    <owlx:Label xml:lang="fr">vin</owlx:Label>
  </owlx:Annotation>
  <owlx:Class owlx:name="&food;PotableLiquid" />
</owlx:Class>
```

Ugyanez RDF szintaktikával:

```
<owl:Class rdf:ID="Wine">
  <rdfs:subClassOf rdf:resource="&food;PotableLiquid"/>
  <rdfs:label xml:lang="en">wine</rdfs:label>
  <rdfs:label xml:lang="fr">vin</rdfs:label>
</owl:Class>
```

## 3. Jogi szövegek

Az előzőleg megismert linked data és szemantikus web gondolatai és eszközei hatékonyan alkalmazhatóak jogi szövegek esetében is. Hiszen ezen dokumentumoknak mindenki számára hozzáférhetőnek és értelmezhetőnek kell lenniük, azonban a hivatalos nyelvezet és a strukturálatlan megjelenési forma ezeket gyakran ellehetetleníti. Egy-egy törvényhez gyakran több oldalas, értelmezést segítő dokumentumokat is kiadnak ennek a javítására.

A törvények és a különböző szervezetek belső működését szabályozó szövegek általában egyszerű szöveggént szerkesztődnek és kerülnek megjelenésre. Erre a probléma egy elterjedőben lévő megoldási javaslat a Metalex szabvány alkalmazása.

### 3.1. Metalex CEN

A Metalex<sup>10</sup> jogi szövegek egységes leírására szolgáló XML formátum. Egységes keretet a tárolásra, logikusan felépített, átlátható és könnyen adaptálható az egyes nemzeti sajátosságokhoz.

Egy rövid lista a fontosabb elemek közül:

- Article
- Category
- Index
- List
- Part
- Sentence
- SentenceFragment
- Subpart
- TextVersion
- Title

Egy rövid lista a fontosabb típusok közül:

- ArticleType
- ListType
- PartType
- ReferringTextType
- TextType

### 3.2. TVSZ

A BME TVSZ-re is igazak a jogi dokumentumokra általánosságban felsoroltak. A következő ábrán látható részleten is megfigyelhetők strukturális jellemzők, azonban ezek nincsenek jelölve.

#### 1. §

##### A Tanulmányi és Vizsgaszabályzat hatálya

(1) A Tanulmányi és Vizsgaszabályzat (a továbbiakban TVSZ) hatálya kiterjed a BME-vel hallgatói jogviszonyban álló, osztatlan, valamint alap- és mesterképzésben, szakirányú továbbképzésben, résziidos képzésben továbbá doktori képzésben részt vevo, tanulóyaikat 2006. szeptember 1. után megkezdett hallgatók tanulmányi és vizsgaügyeire.

(2) A TVSZ egyes rendelkezései az alábbiakra eltéro szabályokat állapítanak meg, amelyeket a szabályzat szövegében külön jelez:

- a.) az alap- (BSc/BA), mester- (MSc/MA) és 2006. szeptember 1. után indult osztatlan képzésekre,
- b.) a 2006. szeptember 1. előtt indult osztatlan képzésekre,
- c.) a doktori képzésre,
- d.) a szakirányú továbbképzésre és minden résziidos képzésre
- e.) az önállóan szervezett, idegen nyelven folyó költségterítéses képzésre.

#### 1. ábra TVSZ részlet

A szövegben a fontosabb részek kiemelése és fogalmak kiemelése kézzel történt, én ennek a folyamatnak csak a végeredményét használtam fel.

### 3.3. EuroVoc

Az EuroVoc<sup>11</sup> az Európai Unió többnyelvű tezauszsa. Az Unió főként parlamenti tevékenységi köreihez kapcsolódó lexikai egységeket fedi le 22 nyelven. Ezen hatalmas méretű tezauszsból az oktatással kapcsolatos szótárt használtam fel magyar és angol nyelven a fogalmak megjelölésére a kiválasztott TVSZ részletekben. A szótár SKOS formátumú.

Részlet egy fogalom angol és magyar megfelelője, valamint szinonimák jelölésével:

<<http://eurovoc.europa.eu/6905>> <<http://www.w3.org/2004/02/skos/core#prefLabel>>  
"dropout"@en .

<<http://eurovoc.europa.eu/6905>> <<http://www.w3.org/2004/02/skos/core#prefLabel>> "iskolai lemorzsolódás"@hu .

<<http://eurovoc.europa.eu/6905>> <<http://www.w3.org/2004/02/skos/core#inScheme>>  
<<http://eurovoc.europa.eu/100217>> .

<<http://eurovoc.europa.eu/6905>> <<http://www.w3.org/2004/02/skos/core#altLabel>>  
"educational dropout"@en .

<<http://eurovoc.europa.eu/6905>> <<http://www.w3.org/2004/02/skos/core#altLabel>> "school dropout"@en .

<<http://eurovoc.europa.eu/6905>> <<http://www.w3.org/2004/02/skos/core#altLabel>>  
"felsőoktatási lemorzsolódás"@hu .

<<http://eurovoc.europa.eu/6905>> <<http://www.w3.org/2004/02/skos/core#altLabel>>  
"iskolaelhagyás"@hu .



## 4. Nyelvi elemzés

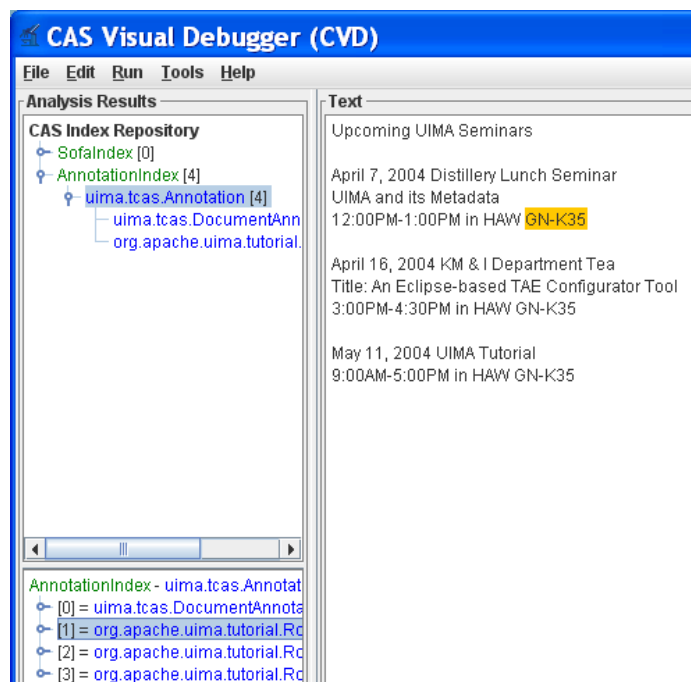
A nagy méretű szövegek számítógéppel történő, automatikus nyelvtani elemzése információkinyerés céljával a mesterséges intelligenciával kapcsolatos kutatások egyik fejlődő ága. Egy ilyen nyelvi elemzés különböző részfeladatokról épül fel melyek eredményei egymásra épülnek, egymásba ágyazhatóak. Ilyen például a szavak töveinek megkeresése, a mondatban elfoglalt pozíció és szerep meghatározása, mondathatárok bejelölése, a rövidítések feloldása.

Ezen feladatok végzi angol nyelvre a Morphadorner<sup>12</sup> eszköz és ennek felhasználásával, magyar nyelvi sajátosságokat figyelembe véve, azokkal kiegészítve készült a Szegedi Tudományegyetem magyarlanc<sup>13</sup> projektje.

### 4.1. UIMA

Az UIMA<sup>14</sup> (Unstructured Information Management Architecture) strukturálatlan szövegek feldolgozásához nyújt keretrendszert. A használatával az elemzési folyamatot részekből építhetjük fel, ezek eredményét külön-külön is nyomon követhetjük, könnyű fejlesztést és más alkalmazásokba történő beépítési lehetőséget biztosít.

Egy egyszerű reguláris kifejezésekre alapuló eszköz futás közben, amely a szövegben a hotelszoba számokat jelöli be:

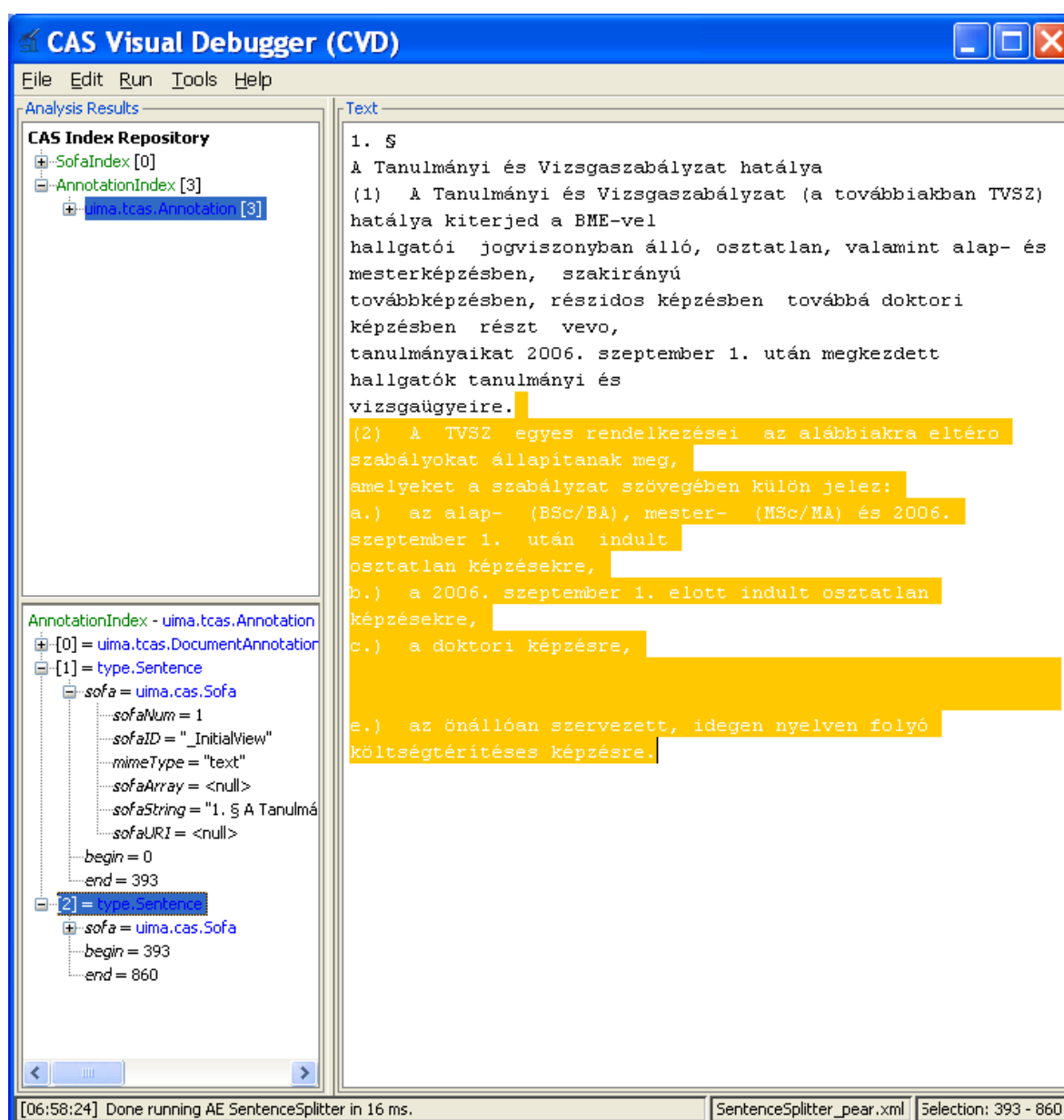


2. ábra UIMA RoomNumberAnnotator<sup>15</sup>

Ehhez a keretrendszerhez tartalmaz különböző modulokat<sup>16</sup> (Sentence Splitter – mondatra bontó, Lemmatizer – szótövesítő, POS tagger – szófaji egyértelműsítő) a magyarlánc projekt is.

#### 4.1.1. Sentence Splitter

A mondatvégi írásjelek keresésével, figyelembe véve a mondatközi előfordulásokat is bejelöli minden szövegben szereplő mondat elejét és végét. A szerkezeti strukturálás és relatív pozíciók meghatározásában hasznos.



3. ábra UIMA Sentence Splitter

#### **4.1.2. Lemmatizer**

A szavak töveinek meghatározás nagyban pontosítja a fogalombejelölést és egyben meghatározza a kapcsolódó toldalékot is amely alapja lehet a következő (POS tagger) elemzőnek.

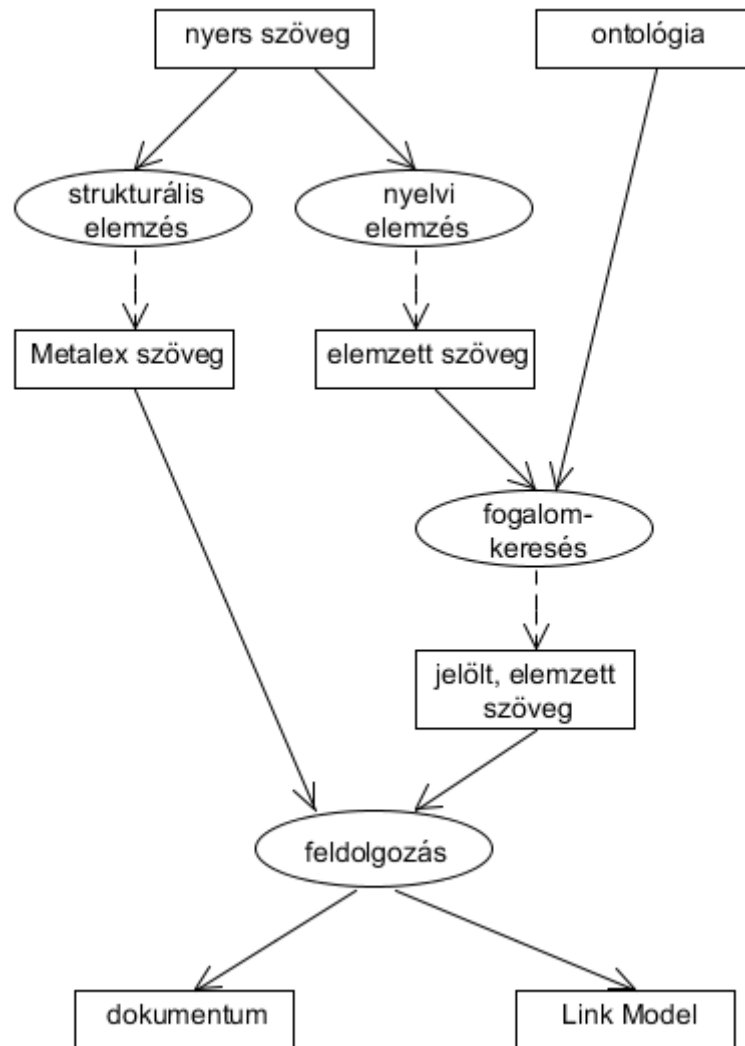
#### **4.1.3. POS tagger**

A szófaji egyértelműsítés<sup>17</sup> szintén szolgálja a fogalombejelölés pontosságának javítását (fogalom és találat szófajának vizsgálata), illetve a kapott MSD<sup>18</sup> kód (Morpho-Syntactic Description – több nyelvre alkalmazható, általános, morfo-szintaktikai jelölésrendszer) további lehetőségeknek, információkinyerésnek az alapja.

## 5. Eredmény

### 5.1. Folyamat

A szöveg feldolgozását a következő folyamat alapján végeztem el:



4. ábra Feldolgozási folyamat

Két bemenet a nyers szöveg és a kézzel készített ontológia. A nyers szövegen párhuzamosan végeztem strukturális és nyelvi elemzést a magyarlánc eszközeinek felhasználásával. Ezen folyamatok kimenete egy Metalex szöveg és egy még nyers, elemzési információkat tartalmazó szöveg.

```
a/Tf, tanulmányi/Áfp-sn, és/Ccsw, vizsgaszabályzat/Nc-sn, hatály/Nc-sn---s3,
/(, 1/N, ), a/Tf, tanulmányi/Áfp-sn, és/Ccsw, vizsgaszabályzat/Nc-sn,
/(, a/Tf, további/Nc-p2, tvsz/X, ), hatály/Nc-sn---s3, kiterjed/Vmip3s---n,
a/Tf, bme/Nc-sn, hallgatói/Áfp-sn, jogviszony/Nc-s2, álló/Áfp-sn,
,,, osztatlan/Áfp-sn, ,,,, valamint/Ccsw, alap-/Oh, és/Ccsw, mesterképzés/Nc-sn,
,,, szakirányú/Áfp-sn, továbbképzés/Nc-sn, ,,,, részidos/X, képzés/Nc-s2, továbbá/Rx,
doktori/Áfp-sn, képzés/Nc-s2, rész/Nc-sa, vevo/X, ,,,, tanulmány/Nc-pa---p3, 2006./Mo,
szeptember/Nc-sn, 1./Ns, után/St, megkezdett/Áfp-sn, hallgató/Nc-pn, tanulmányi/Áfp-sn,
és/Ccsw, vizsgaügyeire./X, /(, 2/N, ), a/Tf, tvsz/X, egyes/Áfp-sn, rendelkezés/Nc-pn---s3,
az/Tf, alábbi/Nc-sn, eltéro/X, szabály/Nc-pa, állapít/Vmip3p---n, meg/Rp, ,,,, amely/Pr3-pa,
a/Tf, szabályzat/Nc-sn, szöveg/Nc-sn, külön/Áfp-sn, jelez/Vmip3s---n, :/:, a./X, ), az/Tf,
alap-/Oh, /(, bsc/X, ///, BÅ/Nc-sn, ), ,,,, mester/Nc-sn, /(, msc/X, ///, ma/Nc-sn, ),
és/Ccsw, 2006./Mo, szeptember/Nc-sn, 1./Ns, után/St, indult/Áfp-sn, osztatlan/Áfp-sn, képzés/Nc-sn,
,,, b./X, ), a/Tf, 2006./Mo, szeptember/Nc-sn, 1./Ns, előtt/Rx, indul/Vmis3s---n,
osztatlan/Áfp-sn, képzés/Nc-sn, ,,,, c./Nc-sn, ), a/Tf, doktori/Áfp-sn, képzés/Nc-ss
,,, d./X, ), a/Tf, szakirányú/Áfp-sn, továbbképzés/Nc-ss, és/Ccsp, minden/Pg3-sn,
részidos/X, képzés/Nc-ss, e./X, ), az/Tf, önálló/Áfp-sw, szervezett/Áfp-sn, ,,,,
idegen/Áfp-sn, nyelv/Nc-sp, folyó/Áfp-sn, költségtérítéssel/Áfp-sn, képzés/Nc-ss, ./.
```

5. ábra Elemzett, nyers szövegrészlet

Az elemzett szövegen végzem el az ontológia fogalmainak keresését és bejelölését, majd az így kapott jelölt, elemzett szöveg és a korábbi Metalex szöveg feldolgozásával készül el a végső „termék”, maga a dokumentum és egy kapcsolódó táblázat, az úgynevezett Link Model.

```
<Title>
  <TextVersion xml:lang="hu">
    A Tanulmányi és <RefTextPart ref="http://eurovoc.europa.eu/925">Vizsgaszabályzat</RefTextPart> hatálya
  </TextVersion>
</Title>
<SubPart id="p1a1s1">
  <IndexDesignation>
    <Index>
      <TextVersion xml:lang="hu">(1)</TextVersion>
    </Index>
  </IndexDesignation>
  <Sentence>
    <TextVersion xml:lang="hu">
      A Tanulmányi és <RefTextPart ref="http://eurovoc.europa.eu/925">Vizsgaszabályzat</RefTextPart> (a továbbiakban TVSZ)
    </TextVersion>
  </Sentence>
</SubPart>
<SubPart id="p1a1s2">
  <IndexDesignation>
    <Index>
      <TextVersion xml:lang="hu">(2)</TextVersion>
    </Index>
  </IndexDesignation>
```

6. ábra „dokumentum”  
Metalex formátumú, jelölt szöveg

## 5.2. Link Model

Az XML formátumú, metaadatokkal, annotációkkal kiegészített dokumentum mellett a folyamat másik kimenete egy táblázat, amely a jelölt fogalmakról tartalmaz információkat. Első megközelítésben tartalmazza a fogalom szövegbeli előfordulását (WORD), a szótövet

(Lemma), az egyértelmű azonosítót (URI), a szövegbeli helyét (Position) és a szófaját (Part of Speech).

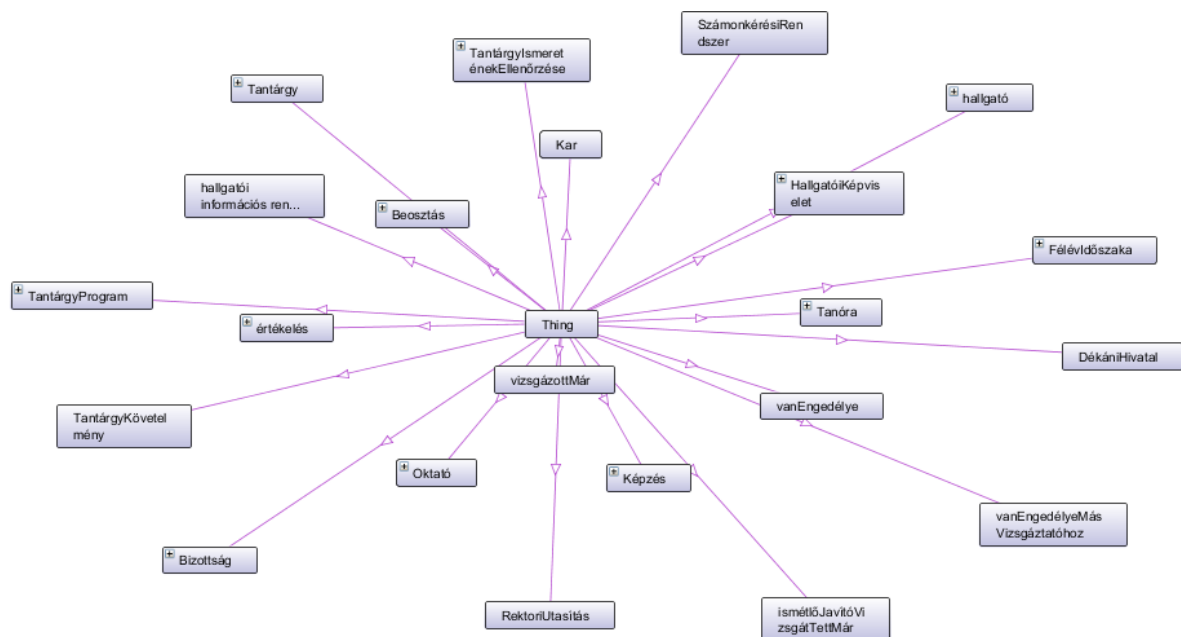
Word	Lemma	URI	Position	Part of Speech
Vizsgaszabályzat	vizsgaszabályzat	http://eurovoc.europa.eu/925	18	noun
vizsgaszabályzat	vizsgaszabályzat	http://eurovoc.europa.eu/925	66	noun
hallgatói	hallgató	http://eurovoc.europa.eu/897	134	adverb

7. ábra Link Model részlet

### 5.3. Emerald WebVisualizer

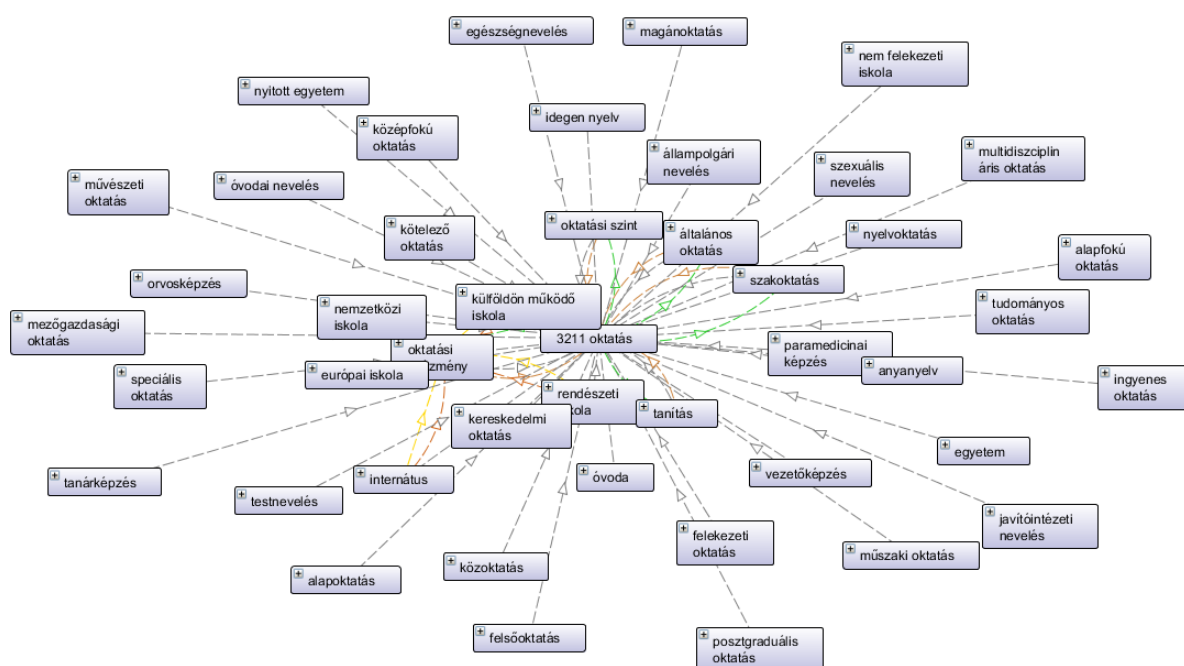
A folyamat eredményeinek megjelenítéséhez a Förhécz András által fejlesztett Emerald WebVisualizer programot használtam. Az eszköz a jogi szövegek megjelenítésére és felhasználóbarát kezelésére, a hozzájuk kapcsolódó terminológiák, a közöttük lévő kapcsolatok szemléltetésére szolgál.

A betöltött és megjelenített TVSZ-es terminológiagráf:



8. ábra TVSZ terminológia

Az EuroVoc Education szótárában szereplő fogalmak gráfja:



9. ábra EuroVoc Education terminológia

A végső eredmény (középen alul az annotált szöveg, felül a kapcsolódó metadatok. Bal oldalt az EuroVoc jobb oldalt pedig a TVSZ fogalomtára és terminológiagráfja):

10. ábra TVSZ az Emerald WebVisualizerben

Az eszköz segítségével a már Metalex szabványt követő XML szöveg, áttekinthető, olvasható formában megjelenik és a fontosabb fogalmak, valamint azok kapcsolata könnyen nyomon követhető.

## **6. Továbbfejlesztési irányok**

Az elkészített folyamat alapját képezheti egy szakértői rendszernek (jelen esetben összekapcsolható a TVSZ-re épített dialógusrendszerrel), több lehetőség is adódik a továbbfejlesztésére.

### **6.1. Pontosság javítás**

A folyamat nagyban automatizált, de a kézi ellenőrzés során mind a fogalom bejelölés, mint a szófaji egyértelműsítés mutat kb. 10%-os hibát. Az elemzők tanításával, pontosításával a jelölési algoritmus (jelenleg szó eleji egyezést keres) javításával érhető el a hibaarány csökkentése.

### **6.2. Elemzés bővítése**

A meglévők mellett új elemzők felvétele, pl. NER (Named Entity Recognition), amely a szövegben szereplő „neves entitásokat”, tulajdonneveket keresi és jelöli be.

### **6.3. Szemantikai tartalom**

A szöveg további szemantikai tartalommal való ellátása, amely a fogalmak közötti kapcsolat, összefüggések feltárását és jelölését jelenti. Ennek alapját képezheti egy fejlesztett Link Model.



## 7. Irodalomjegyzék

---

<sup>1</sup> <https://www.vik.bme.hu/files/00000541.pdf>.

<sup>2</sup> [http://en.wikipedia.org/wiki/World\\_Wide\\_Web](http://en.wikipedia.org/wiki/World_Wide_Web)

<sup>3</sup> <http://linkeddata.org/>

<sup>4</sup> [http://en.wikipedia.org/wiki/Linked\\_data](http://en.wikipedia.org/wiki/Linked_data)

<sup>5</sup> Szeredi Péter, Lukácsy Gergely, Benkő Tamás: A szemantikus világháló elmélete és gyakorlata

<sup>6</sup> <http://www.w3.org/XML/>

<sup>7</sup> <http://www.w3.org/RDF/>

<sup>8</sup> <http://www.w3.org/2004/02/skos/>

<sup>9</sup> <http://www.w3.org/TR/owl2-overview/>

<sup>10</sup> <http://legacy.metalex.eu>

<sup>11</sup> <http://eurovoc.europa.eu>

<sup>12</sup> <http://morphadorner.northwestern.edu/>

<sup>13</sup> <http://www.inf.u-szeged.hu/rgai/nlp?lang=en&page=magyarlanc>

<sup>14</sup> <http://uima.apache.org/>

<sup>15</sup> <http://uima.apache.org/doc-uima-annotator.html>

<sup>16</sup> [http://www.inf.u-szeged.hu/projectdirs/mszny2009/MSZNY2009\\_press\\_b5\\_mod\\_opt.pdf#page=402](http://www.inf.u-szeged.hu/projectdirs/mszny2009/MSZNY2009_press_b5_mod_opt.pdf#page=402)

<sup>17</sup> <http://www.inf.u-szeged.hu/mszny2010/mszny2010.pdf#page=283>

<sup>18</sup> <http://nl.ijs.si/ME/V3/msd/html/>