

Információkinyerés magyar nyelvű gerinc MR leletekből

Kicsi András¹, Pusztai Péter^{1,2}, Szabó Ledényi Klaudia¹, Szabó Endre³,
Berend Gábor¹, Vincze Veronika², Vidács László^{1,2}

¹Szegedi Tudományegyetem, Informatikai Intézet, IKK
Szeged, Árpád tér 2.

²MTA-SZTE Mesterséges Intelligencia Kutatócsoport
Szeged, Tisza Lajos körút 103.

{akicsi,pusztai,ledenyik,berendg,vinczev,lac}@inf.u-szeged.hu

³Szegedi Tudományegyetem
endrebacsi@gmail.com

Kivonat Cikkünkben magyar nyelvű radiológiai leletek automatikus feldolgozásának módszeréről és kezdeti kísérleteink eredményeiről számolunk be. Először bemutatjuk a felhasznált adatbázist és az alkalmazott annotációs elveket, majd ismertetjük kísérleti módszereinket. Bemutatjuk eredményeinket, ezt követően pedig ismertetjük a rendszer jelenlegi erősségeit és gyengébb pontjait, végül szót ejtünk a továbbfejlesztési lehetőségekről is.

Kulcsszavak: radiológia, információkinyerés, nlp, annotáció

1. Bevezető

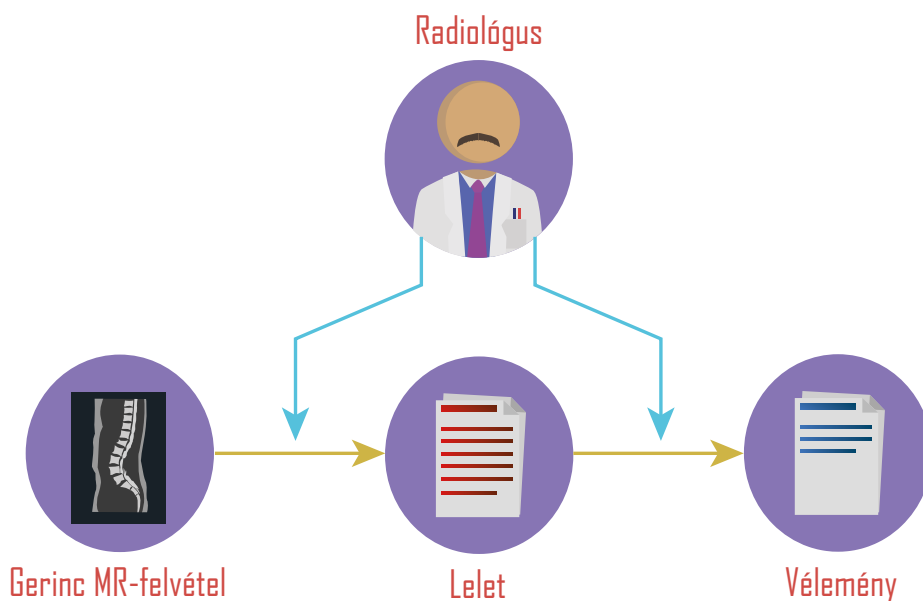
A klinikai gyakorlatban óriási mennyiségű dokumentum keletkezik nap mint nap, melyek között találhatunk leleteket, zárójelentéseket, orvosi lapokban megjelenő publikációkat. Ez a hatalmas szövegmennyiség rengeteg nyers adatot rejt magában, melyek kiaknázásában az informatika egy részterülete, a számítógépes nyelvészet (natural language processing, NLP) nyújthat segítséget. Az információkinyerés feladata, hogy nagy mennyiségű strukturálatlan vagy gyengén strukturált szövegből automatikus eszközökkel összegyűjtse a szövegben meglévő információt. Egy jellemző példa lehet a fehérje–fehérje interakciók kinyerése biológiai szövegekből, ahol a különféle biológiai entitások közti kapcsolatokat kell összegyűjteni. Angol nyelvre már jó ideje léteznek olyan automatizált megoldások, melyek az orvosi szövegekben rejlő információ kiaknázására törekednek, lásd például a páciens dohányzási szokásainak vagy elhízottságának megállapítása orvosi zárójelentések alapján [1,2], a magyar nyelvű orvosi NLP-vel azonban viszonylag kevés munka foglalkozott eddig (lásd pl. [3]).

Ebben a munkában radiológiai leletek automatikus feldolgozásával foglalkozunk. Míg angol nyelvű radiológiai leletek feldolgozására születtek már szép eredmények [4], tudomásunk szerint magyar nyelvű leleteket e szempontból még nem vizsgáltak. Célunk, hogy a leletekből minél több olyan információt nyerjünk

ki automatikus módszerekkel, melyek megkönnyítik a klinikus munkáját. E feladat szakszerű megvalósításához csapatunkban orvos, informatikus és nyelvész kollégák működnek együtt. A továbbiakban részletesen bemutatjuk a felhasznált anyagokat és módszereket, majd ismertetjük elért eredményeinket, végül szót ejtünk a kutatás lehetséges további irányairól is.

2. Motiváció

A radiológiai klinikai gyakorlatban a betegről először elkészül egy modern képalkotó eljárással (például CT, MRI, esetleg röntgen) létrehozott felvétel, melynek alapos vizsgálata során a radiológus szakértő megállapítja az esetleges eltéréseket, rendellenességeket, adott esetben diagnosztizálja a betegséget. Szükség esetén a betegről rendelkezésre álló korábbi leleteket, radiológiai felvételeket is tanulmányozza a minél pontosabb véleményalkotás érdekében. Végül mindezt leletbe foglalja, ahol a fentiek szöveges összegzésén kívül diagnosztikai véleményt kell alkotnia, illetve további vizsgálatokra, terápiára is tehet javaslatot. A munkában egy automatikus diktálórendszer is segíti. A fentieket az 1. ábrán láthatjuk összefoglalva.



1. ábra: A radiológus munkája a vizsgálat után

Cikkünk fő célja, hogy megkönnyítsük a radiológus szakember munkáját. Ennek első lépése lehet a leletek automatikus kivonatolása, azaz kezdeti feladatként

fel kell ismernünk a lelet szövegében található legfontosabb kifejezéseket. Ilyenek lehetnek például az egyes testrészek, elváltozások, betegségek nevei stb., illetve a köztük levő kapcsolatok. Ha ezeket az információkat automatikus úton képesek vagyunk kinyerni, a következő lépcsőben automatikus eszközökkel generálhatunk egy diagnosztikai véleményt, melyet használatkor természetesen a radiológus szakember felülvizsgál, szükség esetén felülbírál. A kutatás jelen fázisában a lelet szövegében található fontos kifejezések minél pontosabb kinyerésére törekszünk: cikkünk további részében ennek folyamatát és jelenlegi eredményességét taglaljuk részletesen.

3. Annotáció

Jelen munkában gerinc MR-leletekkel dolgozunk, melyeket a Szegedi Tudományegyetem Radiológiai Klinikájának munkatársai bocsátottak rendelkezésünkre. A leletekben található személyes jellegű adatokat természetesen az adatvédelmi előírásoknak megfelelően kezeljük vizsgálataink során. Kísérleteinkben 250 lelettel dolgozunk, azonban a későbbiekben várható a leletállomány kibővülése is.

Az adatok megfelelő annotációja a sikeres gépi tanulás alapfeltétele. Ehhez pedig létfontosságú a megfelelően letisztázott alapelvek lefektetése. Ennek érdekében radiológus, informatikus és nyelvész részvételével több találkozáson, iteratíván fejlesztettük ki jelenleg használt annotációs módszerünket. A találkozókra egységesen kijelölt 10 darab leleten végeztek a résztvevők annotációt a korábban megbeszéltek elvek szerint. A tényleges annotáció már az így kialakított elvek alapján történt, a munkához a Brat annotációs eszközt [5] használtuk fel, amelyet a kialakított módszer alapján konfiguráltunk.

A következő címkék jelölésének szükségességét állapítottuk meg:

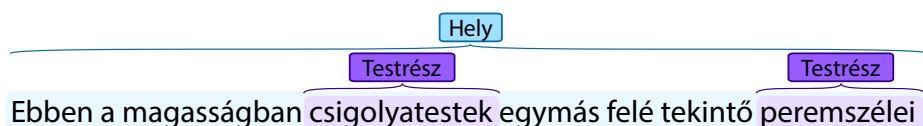
Testrész: Az emberi test egy olyan része, amelyet egy átfogó névvel megnevezhetünk. Például szolgálhatnak rá a 2. ábrán látható kifejezések.



2. ábra: Testrész-ként annotált kifejezések

Hely: Egy helyet ír le az emberi testen belül. Nagy átfedésben van a Testrész címkével, viszont több különböző testrészt és azok viszonyait is felölelheti, az elváltozás helye gyakran ilyen formában van megadva. Példát láthatunk rá a 3. ábrán.

Elváltozás: Az elváltozás, vagy esetleg az elváltozás hiánya, amit a szöveg általában megállapít. Ilyenek például a 4. ábrán látható kifejezések.

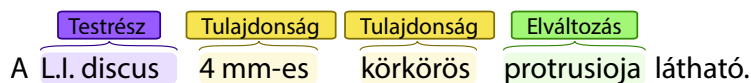


3. ábra: Hely-ként annotált kifejezés



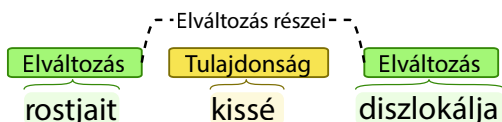
4. ábra: Elváltozás-ként annotált kifejezések

Tulajdonság: Olyan módosító kifejezések, amelyek valamilyen tulajdonságot neveznek meg a leleten belül. Módosíthatják vagy pontosíthatják egy Elváltozás típusát, vagy leírhatják annak fokozatát vagy mértékét is. Mindkettőre láthatunk példát az 5. ábrán látható mondatban.



5. ábra: Egy mondat Tulajdonság-ként annotált kifejezésekkel

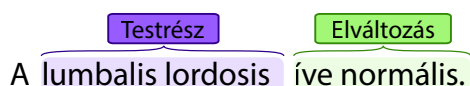
Az entitások közötti relációk közül csupán egyféléét engedtünk meg, ezt *Részei* relációnak hívtuk. Ilyen kapcsolatot azonos címkével ellátott kifejezések között állapítottunk meg, ha úgy találtuk, hogy valójában ugyanazon kifejezés részeit képezik. A reláció fő célja, hogy a kifejezés belsejébe ékelődött egyéb szavak kihagyhatók legyenek a jelölésből. A 6. ábrán látható példa ábrázolja. A feldolgozáskor ezeket a szövegbeli előfordulás sorrendjében kötjük össze és tároljuk egyetlen Elváltozásként.



6. ábra: Részei reláció egy beékelődött Tulajdonság fölött

A találkozók során felmerült kérdések alapján arra jutottunk, hogy a tagadást és bizonytalanságot jelző kifejezéseket nem jelöljük a feladat jelen fázisában, noha a későbbiekben erre mutatkozhat igény. Ezenkívül számos lelet tartalma egy

előző leletre épít, és ehhez viszonyít, mint például „A 2017-es vizsgálatához képest...” típusú mondatokat tartalmazók, amelyek az elváltozásban bekövetkezett változásokra vonatkoznak legnagyobb részben. Ezen leletek automatizált módszerrel kiszűrésre kerültek, ugyanis az anonimizálás után, és mert egyetlen hónap vizsgálatainak leletei álltak csak rendelkezésünkre, a visszakövetés nem megvalósítható. További kérdéses esetet jelentett például, ha az elváltozás a testrész valamilyen tulajdonságára, például szélességére vonatkozik. Itt azt állapítottuk meg, hogy az ilyen, aspektust leíró szavakat az elváltozás részének jelöljük, mint „szélessége normális”, és nem tulajdonságként, mivel a szó szervesen az elváltozáshoz kapcsolódik, valamint nem is minden esetben lehetne tulajdonság címkével sem ellátni. Erre mutat be egy példát a 7. ábrán látható mondat.



7. ábra: Egy testrész valamely aspektusának elváltozása

A jelen cikkben közölt eredményeink a Helyként annotált adatokat nem tartalmazzák, mivel ezek átfedésben lehetnek a Testrészekkel, illetve nagy mennyiségű általános szót tartalmaznak, ezért a kiértékelést jelentősen komplikálnák. Ezek az annotációk azonban a kutatás későbbi fázisaiban szintén nagyon hasznosnak bizonyulnak, így szükségesnek láttuk keresésüket.

4. Kísérletek

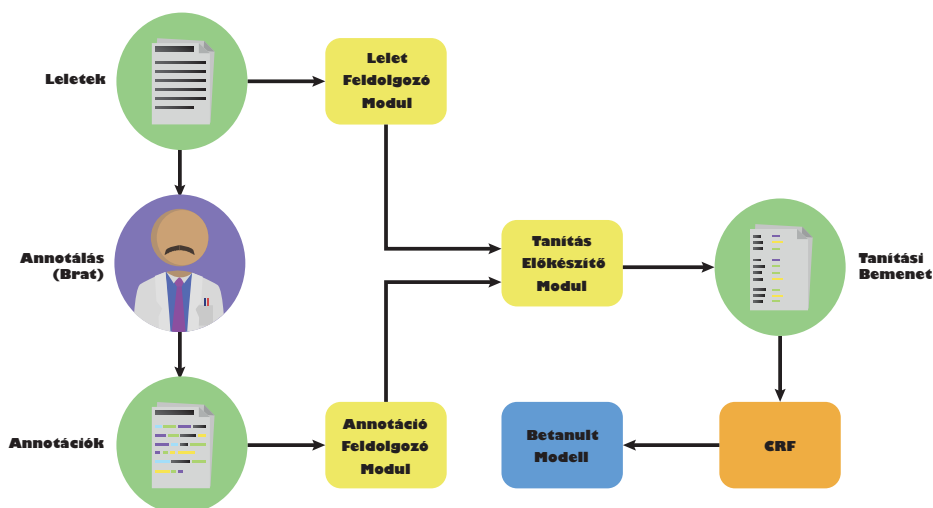
Ebben a fejezetben bemutatjuk a leletekből történő információkinyerést végző szekvenciajelölő osztályozókat, illetve az ennek segítségével elért eredményeinket. A fejezet végén kitérünk a szekvenciajelölő pontosabbá tételének lehetőségeire.

4.1. Gépi tanulás

A teljes folyamat bemutatása a 8. ábrán látható. A leletek feldolgozása után az annotációkat az eredeti leletekhez rendeljük, majd előkészítjük a gépi tanuláshoz az adatokat. Ez lesz az alapja a gépi tanulási szakasznak. A tanuláshoz felhasznált címkék előfordulásainak darabszámát az 1. táblázat szemlélteti.

A testrészek, elváltozások, illetve tulajdonságok beazonosításával kapcsolatban megfogalmazott szekvenciajelölési probléma megoldására egy feltételes valószínűségi modellt (CRF) [6] hoztunk létre. A szekvenciajelölő modell létrehozása, illetve tanítása a CRFsuite [7] csomag segítségével történt.

A szekvenciákban található tokensorozatok leírására az egyes tokenek, illetve a környezetükben – tőlük legfeljebb kettő távolságra – található tokenek felszíni jegyeiből származtattuk a jellemzőkészletet. A jellemzőteret alkotó indikátorváltozók magukat a szóalakokat, illetve a tokenekben előforduló nagybetűket,



8. ábra: A tanulás folyamatának áttekintése

1. táblázat. Az egyes címkék előfordulási gyakorisága a tanító és teszt halmazon. A semelyik érdemi címkével el nem látott tokenek számát az O jelű sor tartalmazza.

Címke	tanító teszt	
Testrészt	3912	983
Elváltozás	5109	1323
Tulajdonság	2106	514
O	7547	1899
Összesen	18674	4719

illetve számjegyeket, kódolták. Az előzőeken túl az egyes tokenek szuffixében álló karakterkettesekből-és hármásokból is alkottunk jellemzőket. A karakterprefixe-kből származtatott jellemzők használatára is tettünk kísérletet, azonban az így kapott modellünk eredményessége elmaradt a kizárólag karaktersuffixeket figyelembe vevőtől. A mondatokra, illetve tokenekre bontást a Spacy [8] nyílt forrású szoftverkönyvtár segítségével végeztük megfelelő konfigurációval.

A jelen kísérletekben a Helyek címkézésére még nem törekedtünk, a Helyként annotált tokeneket az annotálás során érdemi címkével el nem látottakkal azonos módon kezeltük. Az annotációs módszereknél leírt Részei relációk a feldolgozás korai fázisában feloldásra kerülnek.

4.2. Eredmények

A kísérleti eredmények a 2. táblázatban láthatóak. A gépi tanuláshoz a 250 leletből 80%-ot választottunk le tanulóhalmaznak, és a kapott modellt a maradék 20% leleten értékeltük ki. A kiértékelést címkéenként végeztük el, tehát külön

értékeljük ki a *Testrészt*, *Elváltozás* és *Tulajdonság* címkéken mért teljesítményt. Az F1 értékek 90% felett már jó eredményt mutatnak. A modell jó tulajdonsága, hogy kiegyensúlyozott: közel azonos a pontosság (precision) és a fedés (recall) értéke is, egyedül a *Tulajdonság* címke esetén alacsonyabb a fedés. A gerinc egy meglehetősen specifikus területe az emberi testnek, ezért a *Testrészt* címke esetén valószínűleg egy viszonylag kis szókészletre tanulunk, így ebben az esetben a 92,3%-os F1 érték még várhatóan növelhető lesz azáltal, ha például az egyes szóalakok anatómiai atlaszokban való előfordulásának tényét további indikátorváltozók formájában beépítjük a modellünk jellemzőterébe.

2. táblázat. Információkinyerés eredmények CRF szekvenciajelölő használatával

	Pontosság	Fedés	F1-érték	Szupport
Testrészt	0,931	0,916	0,923	983
Elváltozás	0,912	0,904	0,908	1323
Tulajdonság	0,917	0,837	0,875	514
Összesen	0,920	0,896	0,907	2820

Testrészek esetén tipikusnak mondhatóak azok a mondatok, ahol a mondat elején, topik pozícióban szerepel a csigolya, melyről az orvos megállapításokat tesz. Abban az esetben, ha a csigolya rövid jelölése önmagában szerepel (a *csigolya* szó nélkül) és nem kezdő pozícióban, a szekvenciajelölő már hajlamosabb tévedni. Például a következő mondatokban az S.I. csigolyát a modell jelenleg nem ismeri fel Testrészként: *Az L.V. csigolya 3 mm t hátra csúszott az S.I. fölött* *Az L.V. discus víztartalma és magassága csökkent enyhe centralis előbóltosulása a durazsákot eléri.*

A modell szinte egyik lényeges szót sem ismerte fel a következő mondatban: *A L.IV V discus magasságában az anterior longitudinalis szalag vastagabb.* Ezek közül a *vastagabb* szó elváltozásra utal, amit a modell nem ismert fel annak ellenére hogy nehezebb kifejezésekkel is megbirkózik. Szintén gondot jelentett a két latin kifejezés is. Hasonló mondatok esetén felmerül a tanító minta méretének kérdése, itt úgy gondoljuk, hogy a mintaszám növelésével jelentős előrelépés várható.

4.3. Fejlesztési lehetőségek

Bár már az első eredmények is biztatóak, számos továbbfejlesztési lehetőséget látunk a teljes folyamatban. A modell tévesztései alapján és a szakirodalom fényében a következő négy irányvonalat emeljük ki:

Annotáció ellenőrzése A módszer több iteráció során alakult ki, azonban a leletek nagy részét egy orvos annotálta, így szükség van független ellenőrzésre orvosi és nyelvészeti szempontból is.

Nyelvi elemzés Ebben a kísérletben a nyelvi elemzés eredménye még nem jelenik meg a tanulóadatokban. A szófajok és mondatrészek hozzárendelésétől mindenképp mérhető javulást várunk. A tagadás fontos szerepet tölt be a leletek értelmezésénél, melyet az annotálás figyelmen kívül hagy, ezt az információt is a nyelvi elemzés fogja biztosítani. A magyarul elemző [9] tartalmaz dependenciaelemzőt is, melyet szintén használni fogunk a jövőben.

Mély tanulás A jelenlegi gépi tanulási módszer kevés adaton is jól működik. Az adatmennyiség növelésével lehetőség nyílik mély neuronhálókat használatára, melyet a gépi tanulás sok területén sikerrel alkalmaztak a közelmúltban.

További leletek További leletek annotálásával tovább pontosítható a gépi tanulás. Nagyobb mennyiségű lelet további annotálása korlátokba ütközik, viszont a leletek önmagukban is fontos információt jelentenek a mély tanulás számára.

5. Kapcsolódó kutatások

A radiológiai leletek jellemzően még mindig szabad megfogalmazású, a radiológus által diktált szövegek, nem pedig jól kategorizált adatgyűjtemények, így azok információtartalmának megfelelő kinyerése kihívás elé állítja a kutatókat. Általánosságban elmondhatjuk, hogy a jelenleg vezető kutatások az adatok kinyerésére valamilyen, gyakran gépi tanulóval kiegészített, természetesnyelv-feldolgozó (NLP) módszert használnak. Az alkalmazások köre a kinyert információ típusától függően széles spektrumon változik. Többek között beszélhetünk diagnózissegéd [10], [11], [12], diagnosztikai minőségbiztosítást [13], [14], [15], [16], a leletek automatikus BNO kódolását végző [17], a nem várt elváltozásokra adott válaszlépéseket [18], vagy a további vizsgálatokra vonatkozó ajánlásokat figyelő [19], illetve a páciens egészségi állapotát nyomon követő alkalmazásokról [20]. A közelmúltban több olyan összefoglaló cikk is megjelent, mely jól bemutatja az elmúlt egy évtizedben történt fontosabb előrelépéseket [21], [22], [23], [24], [25], [26].

A terület folyamatos bővülése ellenére a nemzetközi szakirodalom viszonylag szegényesnek mondható a kifejezetten gerincröntgen leleteket, NLP és gépi tanulóval feldolgozó tanulmányokat illetően. Tan és munkatársai egy szabályalapú és egy gépi tanulóval alapuló rendszer teljesítményét hasonlították össze 26 alsóháti fájdalomra utaló orvosi megállapítás radiológiai leletekben történő felismerésében [27]. A feladatot a gépi tanulóval alkalmazó rendszer 0,98, míg a szabályalapú rendszer 0,90 AUC (vevő működési karakterisztika görbe alatti terület) érték mellett teljesítette. A szerzők egy másik, 2018-as tanulmányukban reguláris kifejezéseket alkalmazó, szabályalapú NLP algoritmust mutattak be, mely a radiológiai leletekben az 1-es típusú Modic véglemez elváltozásokat 0,79 F1-érték mellett ismerte fel [28]. A szerzők a reguláris kifejezéseken alapuló algoritmus hátrányaként említik, hogy a gerincleletek szövegének változatossága miatt nehéz minden esetet lefedő szabályrendszert kialakítani. Wang és szerzőtársai 6 általános, csonttrikulásból fakadó töréstípus szabad megfogalmazású radiológiai leletekben történő felismerésére fejlesztettek szabályalapú NLP alkalmazást,

mely elsősorban reguláris kifejezések használatával végzi az osztályozást [29]. A gerinctörésem esetek felismerésében a modell 0,91 F1-értéket mutatott. Hassanpour és munkatársai SVM technológiával kiegészített CRF modellt fejlesztettek az orvosi szempontból releváns elváltozások radiológiai leletekben történő felismerésére. Az alkalmazás ezen túlmenően vizsgálta az elváltozások állapotában bekövetkező változások mértékét és jelentőségét is [30]. A modell a jelentős elváltozások azonosításában 0,75, míg a változás mértékének azonosításában 0,95 F1-értéket ért el. Xu és szerzőtársai gyakori szöveges mintázatok bányászatával (labeled sequential pattern, LSP) támogatott CRF modellt fejlesztettek a radiológiai leletekben található további vizsgálatokat javasoló mondatok felismerésére [31]. Az LSP feladata az ajánlást nagy valószínűséggel nem tartalmazó mondatok kiszűrése volt, míg a CRF az ajánlást tartalmazó mondatok azonosítását végezte. A modell az ajánlást tartalmazó mondatok felismerésében 0,88 F1-értéket ért el.

Magyar nyelvű klinikai szövegek feldolgozásában is születtek már eredmények [32]. Például orvosi szövegek automatikus szegmentálására és az orvosi rövidítések automatikus kezelésére gépi tanuláson alapuló rendszert fejlesztettek [33,34]. A leletek szövegeiben gyakran előforduló elírások, elgépelések automatikus javítása szintén elengedhetetlen egy jól működő információkinyerő alkalmazáshoz [35]. Az utóbbi években ezen felül megvalósult egy szemészeti klinikai keresőrendszer [36], illetve egy magyar nyelvű orvosi leletekben, releváns kifejezések azonosítására specializált, nem felügyelt módszereket alkalmazó rendszer is [37].

Tanulmányunk egy további lépés a magyar nyelvű szövegekből történő klinikai információkinyerés felé, mely egyelőre a gerincröntgen leletekben található elváltozás, tulajdonság és testrész típusú szavak detektálásra szorítkozik. Noha összehasonlítható eredmények magyar nyelvre egyelőre nem állnak rendelkezésre, azonban a nemzetközi szakirodalomban található számszerű eredményekkel összevetve rendszerünk versenyképesnek tűnik.

6. Összegzés

Cikkünkben magyar nyelvű radiológiai leletek automatikus feldolgozásának módszeréről és az első kapcsolódó kísérletekről számoltunk be. Kidolgoztunk egy módszert a leletek legfontosabb elemeinek annotálására, melyet radiológus, nyelvész és informatikus részvételével iteratíván, több körben finomítottunk. A módszer alapján a radiológus 250 gerinc MR lelet annotálását végezte el. Az információkinyeréshez gépi tanulást alkalmaztunk, szem előtt tartva, hogy a jövőben a gerinc MR-felvételeken kívül tágabb felhasználási területeket is meg szeretnénk nyitni. Egy CRF tanulót alkalmaztunk, mely egy korszerű és gyakorta használt megoldás névelem-felismerésre. Az első kísérletek 87,5%-92,3% közötti F1-értékkel zárultak, melyek összemérhetőek a hasonló célokat kitűző, de angol nyelvű leleteken elvégzett kísérletekkel. A jelenlegi megoldás számos fejlesztési lehetőséget kínál, mint a nyelvi elemzés felhasználása és a mély tanulás alkalmazása.

Köszönetnyilvánítás

A jelen cikkben közölt kutatást részben az Emberi Erőforrások Minisztériuma támogatta (20391-3/2018/FEKUSTRAT).

Hivatkozások

1. Uzuner, O., Goldstein, I., Luo, Y., Kohane, I.: Identifying Patient Smoking Status From Medical Discharge Records. *Journal of the American Medical Informatics Association: JAMIA* **15**(1) (2008) 14–24
2. Uzuner, O.: Recognizing Obesity and Comorbidities in Sparse Data. *Journal of the American Medical Informatics Association* **16**(4) (2009) 561–570
3. Orosz, Gy., Novák, A., Prószéky, G.: Lessons Learned from Tagging Clinical Hungarian. *International Journal of Computational Linguistics and Applications* **5** (2014)
4. Friedman, C., Alderson, P.O., Austin, J.H.M., Cimino, J.J., Johnson, S.B.: A General Natural-language Text Processor for Clinical Radiology. *Journal of the American Medical Informatics Association* **1**(2) (1994) 161–174
5. Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., Tsujii, J.: brat: A Web-based Tool for NLP-Assisted Text Annotation. In: *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France, Association for Computational Linguistics* (2012) 102–107
6. Lafferty, J.D., McCallum, A., Pereira, F.C.N.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: *Proceedings of the Eighteenth International Conference on Machine Learning. ICML '01, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc.* (2001) 282–289
7. Okazaki, N.: CRFsuite: A Fast Implementation of Conditional Random Fields (CRFs). <http://www.chokkan.org/software/crfsuite/> (2007)
8. Honnibal, M.: spaCy: Industrial-Strength Natural Language Processing. <https://spacy.io/> (Utoljára látogatva: 2018-11-10)
9. Zsibrita, J., Vincze, V., Farkas, R.: magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian. <http://rgai.inf.u-szeged.hu/index.php?lang=en&page=magyarlanc> (2013)
10. Pham, A.D., Névél, A., Lavergne, T., Yasunaga, D., Clément, O., Meyer, G., Morello, R., Burgun, A.: Natural Language Processing of Radiology Reports for the Detection of Thromboembolic Diseases and Clinically Relevant Incidental Findings. *BMC Bioinformatics* **15**(1) (2014) 266
11. Rink, B., Roberts, K., Harabagiu, S., Scheuermann, R.H., Toomay, S., Browning, T., Bosler, T., Peshock, R.: Extracting Actionable Findings of Appendicitis from Radiology Reports Using Natural Language Processing. *AMIA Joint Summits on Translational Science Proceedings. AMIA Joint Summits on Translational Science* **2013** (2013) 221
12. Solti, I., Cooke, C.R., Xia, F., Wurfel, M.M.: Automated Classification of Radiology Reports for Acute Lung Injury: Comparison of Keyword and Machine Learning Based Natural Language Processing Approaches. In: *Proceedings - 2009 IEEE International Conference on Bioinformatics and Biomedicine Workshops, BIBMW 2009. Volume 2009., NIH Public Access* (2009) 314–319

13. Raja, A.S., Ip, I.K., Prevedello, L.M., Sodickson, A.D., Farkas, C., Zane, R.D., Hanson, R., Goldhaber, S.Z., Gill, R.R., Khorasani, R.: Effect of Computerized Clinical Decision Support on the Use and Yield of CT Pulmonary Angiography in the Emergency Department. *Radiology* **262**(2) (2012) 468–474
14. Ip, I.K., Mortele, K.J., Prevedello, L.M., Khorasani, R.: Focal Cystic Pancreatic Lesions: Assessing Variation in Radiologists' Management Recommendations. *Radiology* **259**(1) (2011) 136–41
15. Siström, C.L., Dreyer, K.J., Dang, P.P., Weilburg, J.B., Boland, G.W., Rosenthal, D.I., Thrall, J.H.: Recommendations for Additional Imaging in Radiology Reports: Multifactorial Analysis of 5.9 Million Examinations. *Radiology* **253**(2) (2009) 453–61
16. Dang, P.A., Kalra, M.K., Blake, M.A., Schultz, T.J., Stout, M., Lemay, P.R., Freshman, D.J., Halpern, E.F., Dreyer, K.J.: Natural Language Processing Using Online Analytic Processing for Assessing Recommendations in Radiology Reports. *Journal of the American College of Radiology* **5**(3) (2008) 197–204
17. Farkas, R., Szarvas, Gy.: Eljárás radiológiai leletek automatikus BNO kódolására. In Tanács, A., Csendes, D., eds.: V. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2007), Szeged, Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szegedi Tudományegyetem Informatikai Tanszékcsoport (2007) 149–157
18. Dutta, S., Long, W.J., Brown, D.F., Reisner, A.T.: Automated Detection Using Natural Language Processing of Radiologists Recommendations for Additional Imaging of Incidental Findings. *Annals of Emergency Medicine* **62**(2) (2013) 162–169
19. Yetisgen-Yildiz, M., Gunn, M.L., Xia, F., Payne, T.H.: Automatic Identification of Critical Follow-Up Recommendation Sentences in Radiology Reports. *AMIA ... Annual Symposium Proceedings / AMIA Symposium. AMIA Symposium 2011* (2011) 1593–602
20. Cheng, L.T.E., Zheng, J., Savova, G.K., Erickson, B.J.: Discerning Tumor Status from Unstructured MRI Reports-Completeness of Information in Existing Reports and Utility of Automated Natural Language Processing. *Journal of Digital Imaging* **23**(2) (2010) 119–132
21. Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., Liu, S., Zeng, Y., Mehrabi, S., Sohn, S., Liu, H.: Clinical Information Extraction Applications: A Literature Review (2018)
22. Pons, E., Braun, L.M., Hunink, M.G., Kors, J.A.: Natural Language Processing in Radiology: A Systematic Review. *Radiology* **279**(2) (2016) 329–343
23. Ford, E., Carroll, J.A., Smith, H.E., Scott, D., Cassell, J.A.: Extracting Information from the Text of Electronic Medical Records to Improve Case Detection: A Systematic Review. *Journal of the American Medical Informatics Association* **23**(5) (2016) 1007–1015
24. Cai, T., Giannopoulos, A.A., Yu, S., Kelil, T., Ripley, B., Kumamaru, K.K., Rybicki, F.J., Mitsouras, D.: Natural Language Processing Technologies in Radiology Research and Clinical Applications. *RadioGraphics* **36**(1) (2016) 176–191
25. Yim, W.w., Yetisgen, M., Harris, W.P., Kwan, S.W.: Natural Language Processing in Oncology. *JAMA Oncology* **2**(6) (2016) 797
26. Meystre, S.M., Savova, G.K., Kipper-Schuler, K.C., Hurdle, J.F.: Extracting Information from Textual Documents in the Electronic Health Record: A Review of Recent Research. *Yearbook of Medical Informatics* (2008) 128–44
27. Tan, W.K., Hassanpour, S., Heagerty, P.J., Rundell, S.D., Suri, P., Huhdanpaa, H.T., James, K., Carrell, D.S., Langlotz, C.P., Organ, N.L., Meier, E.N., Sherman, K.J., Kallmes, D.F., Luetmer, P.H., Griffith, B., Nerenz, D.R., Jarvik, J.G.:

- Comparison of Natural Language Processing Rules-Based and Machine-Learning Systems to Identify Lumbar Spine Imaging Findings Related to Low Back Pain (2018)
28. Huhdanpaa, H.T., Tan, W.K., Rundell, S.D., Suri, P., Chokshi, F.H., Comstock, B.A., Heagerty, P.J., James, K.T., Avins, A.L., Nedeljkovic, S.S., Nerenz, D.R., Kallmes, D.F., Luetmer, P.H., Sherman, K.J., Organ, N.L., Griffith, B., Langlotz, C.P., Carrell, D., Hassanpour, S., Jarvik, J.G.: Using Natural Language Processing of Free-Text Radiology Reports to Identify Type 1 Modic Endplate Changes. *Journal of Digital Imaging* **31**(1) (2018) 84–90
 29. Wang, Y., Mehrabi, S., Sohn, S., Atkinson, E., Amin, S., Liu, H.: Automatic Extraction of Major Osteoporotic Fractures from Radiology Reports using Natural Language Processing. In: *Proceedings - 2018 IEEE International Conference on Healthcare Informatics Workshops, ICHI-W 2018, IEEE* (2018) 64–65
 30. Hassanpour, S., Bay, G., Langlotz, C.P.: Characterization of Change and Significance for Clinical Findings in Radiology Reports Through Natural Language Processing. *Journal of Digital Imaging* **30**(3) (2017) 314–322
 31. Xu, Y., Tsujii, J., Chang, E.I.C.: Named Entity Recognition of Follow-Up and Time Information in 20 000 Radiology Reports. *Journal of the American Medical Informatics Association* **19**(5) (2012) 792–799
 32. Siklósi, B., Novák, A.: A Magyar Beteg. In Tanács, A., Varga, V., Vincze, V., eds.: *X. Magyar Számítógépes Nyelvészeti Konferencia, Szeged, Szegedi Tudományegyetem, Informatikai Tanszékcsoport* (2014) 188–198
 33. Orosz, Gy., Novák, A., Prószéky, G.: Hybrid Text Segmentation for Hungarian Clinical Records. In Castro, F., Gelbukh, A., González, M., eds.: *Advances in Artificial Intelligence and Its Applications: 12th Mexican International Conference on Artificial Intelligence, MICAI 2013, Mexico City, Mexico, November 24-30, 2013, Proceedings, Part I, Berlin, Heidelberg, Springer Berlin Heidelberg, Springer Berlin Heidelberg* (2013) 306–317
 34. Siklósi, B., Novák, A.: Rec. et exp. aut. Abbr. mnyelv. KLIN. szövb-en – Rövidítések Automatikus Felismerése és Feloldása Magyar Nyelvű Klinikai Szövegekben. In Tanács, A., Varga, V., Vincze, V., eds.: *X. Magyar Számítógépes Nyelvészeti Konferencia, Szeged, Szegedi Tudományegyetem, Informatikai Tanszékcsoport* (2014) 167–176
 35. Siklósi, B., Orosz, Gy., Novák, A., Prószéky, G.: Automatic Structuring and Correction Suggestion System for Hungarian Clinical Records. In De Pauw, G., de Schryver, G.M., Forcada, M.L., M. Tyers, F., Waiganjo Wagacha, P., eds.: *8th SaLTMiL Workshop on Creation and Use of Basic Lexical Resources for Less-Resourced Languages, Istanbul* (2012) 29–34
 36. Siklósi, B., Novák, A.: Digitális Konzílium – Egy Szemészeti Klinikai Keresőrendszer. In Tanács, A., Varga, V., Vincze, V., eds.: *XII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2016), Szeged, Szegedi Tudományegyetem, Szegedi Tudományegyetem* (2016) 230–240
 37. Siklósi, B., Novák, A.: Nem Felügyelt Módszerek Alkalmazása Releváns Kifejezések Azonosítására és Csoportosítására Klinikai Dokumentumokban. In Tanács, A., Varga, V., Vincze, V., eds.: *XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015), Szeged, Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szegedi Tudományegyetem Informatikai Tanszékcsoport* (2015) 237–248