

# Klinikai leletek strukturálása mondatszintű címkézéssel

Szabó Ledenyi Klaudia<sup>1</sup>, Pusztai Ágnes<sup>1</sup>, Kicsi András<sup>1</sup>, Vidács László<sup>1</sup>

<sup>1</sup>Szegedi Tudományegyetem, Szoftverfejlesztés Tanszék  
Szeged, Dugonics tér 13.  
{ledenyik,pusztai,akicsi,lac}@inf.u-szeged.hu

**Kivonat** A szöveges orvosi leletek különböző jellegű információkat tartalmaznak, mint például a vizsgálati technika, a páciens panaszai, az orvos által megállapított elváltozások, vagy az orvos véleménye. Az információ strukturálása nagyban megkönnyítheti a leletek értelmezését, adminisztrációját, és gépi feldolgozását is. Cikkünkben egy huBERT modellel alapuló megoldással címkézzük fel magyar nyelvű gerinceletek szövegét mondatonként. A modell egy egyszerű utófeldolgozás után 98% fölötti pontossági értéket produkált. Eredményeink alapján a modell jól általánosít más anatómiai régiók (97%), és más intézmények (93%) leleteire is, jó alternatívát nyújtva az anatómiai régiókra vagy intézményekre optimalizált szabályalapú módszerek helyett.

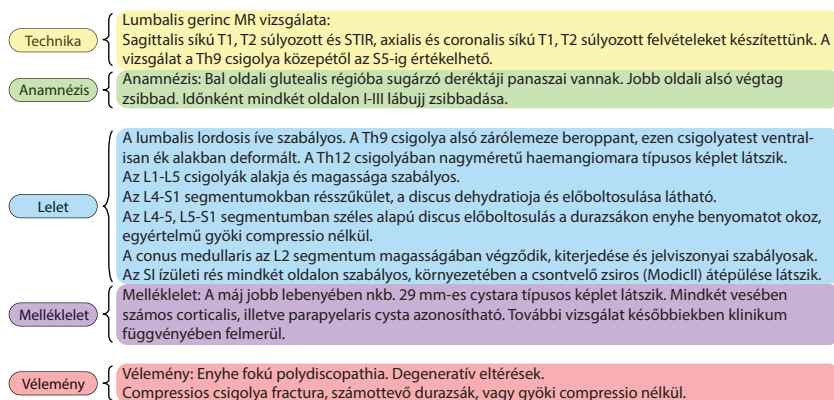
**Kulcsszavak:** leletezés, lelet-strukturálás, mondat címkézés, nlp

## 1. Bevezetés

Az egészségügyi vizsgálatok során számos szöveges leírás keletkezik, amellyel a szakemberek leírják a páciens panaszait, korábbi klinikai információit, az orvosok által kért vizsgálatokat, a vizsgálatok során tapasztalt rendellenességeket, az orvosok véleményét, a javasolt kezeléseket, vagy egyéb más adatot a páciensről. A leletek formátumában gyakran vannak megkötések, a leletek struktúrájának fenntartására irányuló törekvések, ám ezek formája intézményenként, alkalmazása pedig orvosonként változhat. Ha van is ilyen kikényszerített és megfelelően fenntartott struktúra, az ilyen leíró mezők szövege akkor is általában kötetlen természetesnyelvű szöveg.

A strukturált leletezés az irodalom szerint nagyobb pontosságot és konzisztenciát eredményez az orvosok részéről, és a leleteket olvasó klinikusok is nagyobb elégedettséget és jobb értelmezési képességet mutattak (Bosmans és mtsai (2011), Ganeshan és mtsai (2018)). A leletek gépi feldolgozásához is nagyban hozzájárulhat az, ha legalább az adott dokumentáció típusát meg tudjuk különböztetni, hiszen például a vizsgálati technika, vagy a beteg korábbi vizsgálatainak leírásához kifejezetten más célzott módszerek lehetnek szükségesek, mint a tényleges leletszöveg, vagy vélemény elemzéséhez.

Magyarországon az intézmények többségében a lelet formátuma nem teljesen kötött, az információ többségét általában egy szövegdobozban rögzítik. Ezen belül ugyan a különböző szakaszokat gyakran tagolni szokták bekezdésekkel, de ez



1. ábra: Egy példalelet, releváns szakaszaival címkézve

egyáltalán nem kötelező, és egy-egy szakasz megléte is opcionális, nem feltétlen kötött a sorrendjük, illetve egy szakasz több bekezdésből is állhat. Számos intézményben a szakaszt felcímkézik a szövegben, de ezek a címkék akár intézményen belül is változhatnak, és szintén nem kötelezőek. Egy példaleletet láthatunk az 1. ábrán, amely a későbbiekben általunk is tárgyalt szakaszokat szemlélteti. Kutatásunk során radiológiai leletekhez volt hozzáférésünk, így a továbbiakban ezen területet fogjuk részletesebben vizsgálni.

Tekintsük át röviden a példa szakaszait: Az első szakaszban a vizsgálati technika leírását látjuk. Ez megadja, hogy milyen vizsgálat készült, mi az, ami alapján a lelet le fogja írni a megállapításokat. A vizsgálatok általában az érintett testrészek szempontjából is fontosak. Szintén gyakori azonban, hogy több vizsgálat eredménye is bekerül egy leletbe, ilyenkor az új vizsgálat vizsgálati technikáját is megadják, a lelet egy későbbi bekezdésében. Az anamnézis szakasz a páciens előéletét adja meg, panaszait írja le, illetve korábbi vizsgálatait vagy kezelését. Ezt a szakaszt jelentősen nagyobb szókinccs jellemzi, hiszen tartalma kevésbé szakmai (például „három hónappal korábban elesett a kádban”). A részletes lelet leírása képezi a tartalom legfontosabb részét, ebben az orvos részletesen leírja a vizsgálat során, vagy a felvételeken látott összes tapasztalatát, akár a korábbi felvételekhez is viszonyítva. Általában nem csak a kóros eltéréseket jegyzi meg, hanem azt is, ha valamit a normálisnak megfelelő állapotúnak lát. A leletekben szerepelhet melléklet is, ez olyankor releváns, ha a felvételeken a specialista olyan kóros elváltozást is lát, amely ugyan nem szigorúan része a vizsgálatnak, de mégis fontos megjegyezni, hiszen a páciensnek ugyanúgy problémát okozhat, mint a célzott területen lévő elváltozások. A leletet általában vélemény zárja, amely szintén az orvos tapasztalatain alapszik, ez szinte minden esetben a lelet által már leírt információt tartalmazza, ám gyakran szűkszavúbb és szakmaibb megfogalmazásban, illetve a lelet lényeges tartalmát foglalja össze. Szintén itt szoktak megjelenni az orvos javaslatai is a további kezelésre, esetleg újabb vizsgálatokra. A leírtakból valójában csak a részletes lelet a kötelező, de a technika és vélemény jelenléte szintén igen gyakori.

A szakaszok megfelelő elkülönítése különösen érdekes a későbbi automatizált feldolgozás szempontjából. A szakaszok különböző információt tartalmaznak, ennek megfelelő kinyerése hozzájárulhat a jobb és átláthatóbb adminisztrációhoz, és minőségibb leletezéshez. A leletszöveg és vélemény például egymáshoz képest redundanciát is tartalmaz, könnyű belátni, hogy ez hogyan eredményezhet később zajt. A páciensek szemszögéből valós alkalmazás lehet a leletek megértését célzó elektronikus lelet, ehhez szintén szükség van a lelet megfelelő tagolására. Ilyen esetben értékes funkciót jelenthet az is, ha nyomtatott leletet is feltölthetnek, ehhez viszont kritikus lehet a tartalmat megkülönböztetni a papíron található többi metaadattól, mint az intézmény, vagy a páciens adatai.

Jelen írásunkhoz valós leleteken végeztünk mondat szintű annotációt a szakaszok azonosításával. Fókuszunkban radiológiai gerincclepek voltak, a tanítást ezeken végeztük, egy intézmény leletein. Felmerül azonban a kérdés, hogy mennyire lehet általánosítható az így tanított modell másik intézményből származó, vagy az emberi test egy másik területét érintő leletek esetére. Kísérleteinkben ezért a következő három kutatási kérdést fogjuk vizsgálni:

- **1. kérdés:** Képes-e egy gépi tanulási modell pusztán a leletszövegből – bekezdések és egyéb címkek megkövetelése nélkül – megfelelően szakaszokra bontani a leleteket?
- **2. kérdés:** Egy terület és intézmény leletein tanított modell hogyan teljesít más területről/intézményekből származó adatok strukturálásában?
- **3. kérdés:** Hogyan teljesít az általános módszer az intézményenkénti szabályalapú megközelítésekkel szemben?

## 2. Háttér

A szabadszöveges leletek feldolgozását nagyban elősegítik a természetesnyelv-feldolgozó (NLP) módszerek. Ezek fejlődésével, a különböző nyelvfeldolgozási feladatokra használt, kezdeti szabályalapú és heurisztikus módszereket a szó szintű reprezentációk váltották fel. A reprezentációk létrehozásának fejlődésével azok egyre kifinomultabbá váltak, és egyre több információt foglaltak magukba. A kezdeti reprezentációk, mint például a word2vec (Mikolov és mtsai (2013)), gloVe (Pennington és mtsai (2014)) és a fastText (Joulin és mtsai (2016)) a szavak egyéni, karakterszinű tulajdonságait használták fel, azonban a szövegkörnyezet által rejtett információkat nem tartalmazták. A szövegkörnyezeti információk felhasználására az ELMo (Peters és mtsai (2018)) és BERT modellek (Devlin és mtsai (2019)) nyújtottak megoldást.

A nyelvfeldolgozás területén történő fejlődés az orvosi leletek szemantikus szekciókra való felbontását (section segmentation) is pozitívan befolyásolta. A kezdeti kísérletekben statisztikai és gépi tanuló módszereket alkalmaztak. A statisztikai módszerekben a kézzel megtervezett jellemzők játszottak nagy szerepet, köztük az olyan szóalakjellemzők, mint a nagy kezdőbetűk, számok, üres sorok jelenléte és pozíciója. Taira és társai publikációjukban (Taira és mtsai (2001)) a

szabályalapú stukturális elemzés után, több lépésben történt a szekciók címkéinek meghatározása. Egy későbbi kutatás szerzői a heurisztikus határazonosítók után egy support vector machine (SVM) modellel próbálták meghatározni az egyes szekciókhoz tartozó címkéket (Apostolova és mtsai (2009)). Tepper és társai (Tepper és mtsai (2012)) a mondatszintű klasszifikáció helyett sorszinten dolgoztak, a címkék meghatározása pedig a MaxEnt elnevezésű modellel történt. Publikációjukból kiderült, hogy a sorszintű azonosítás hátránya, hogy a külön bekezdések hiánya negatívan befolyásolja az eredményt.

Egy 2019-es áttekintő kutatás (Pomares-Quimbaya és mtsai (2019)) szerint az elektronikus orvosi leletekben (EHR) a statisztikai és szabályalapú módszerek domináltak, és a legjobb eredményeket ezek kombinációjával tudták elérni. Ezek az eredmények azonban még mindig nagyban függtek a kiépített kézi jellemzők minőségétől. A statisztikai és szabályalapú módszerek időigényesek, és nem általánosíthatók, ami további megoldások keresését vonta maga után.

A szóbeágyazások és neurális hálók elterjedésével a szekciók szegmentációjában is megjelentek az ezzel kapcsolatos kutatások (Sadoughi és mtsai (2018)). Az ELMo és a BERT sikere után Rosenthal és munkatársai (Rosenthal és mtsai (2019)) 400 CTPA (Computed Tomography Pulmonary Angiography) lelet alapján, összesen több mint 4 000 mondat felhasználásával készítették mondatszintű, szöveggörnyezetet is figyelembe vevő, terület-specifikus modellt. Előre adott szekciók klasszifikációjában a szerzők egy RNN modellel a terület-specifikus leleteken 93%-os F1-mértékkel sikeresen klasszifikálnak, és a BERT ezt a teljesítményt majdnem tökéletes, 99%-ra emeli. Azonban a gyakorlatban ezek a szekciók nem előre adóttak, így ismeretlen szekcióhatárokkal, mondatklasszifikációval is bemutatják a szerzők a BERT fölényét: a 63%-os RNN modell eredményét egy 74%-os BERT modellel javítják.

Kuling és társai elkészítették a BI-RADS BERT-et (Kuling és mtsai (2022)), mely egy 155 000 radiológiai leleten előtanított BERT modell. Az előtanított modellt többek között szakasz-szegmentációs feladatra is finomhangolták, úgy, hogy a BERT reprezentációit egy 3-rétegű teljesen összekötött és egy Tanh aktivációs függvényvel ellátott hálóval egészítették ki. A második háló kézzel szerkesztett jellemzők (az előző mondat klasszifikációja, a mondat sorszáma, és a leletben található összes mondat száma) alapján egy 128 elemű kiegészítő jellemzővektort készítettek. Az egyéb jellemzőket tartalmazó vektor BERT reprezentációhoz adása több százalékos javulást eredményezett. Az előtanított BI-RADS BERT további 900 leleten való szekció szegmentációra való finomhangolás után 89,5%-os F1, és 97,8%-os pontosságú eredményt mutatott. A szerzők által bemutatott kiegészítő adatokon tanult hálóval ezt az F1-mértéket 94,8%-ra emelték, 94,1%-os pontossággal.

A magyar nyelvű leletek tekintetében a szakaszokra bontás egyelőre nem kapott sok figyelmet. Ugyan az orvosi leletekkel foglalkozó kutatások már elterjedtek (Siklósi és Novák (2014), Kicsi és mtsai (2019)), ezek általában csak a cikkünkben részletes leletnek vagy véleménynek nevezett szakaszokkal foglalkoznak. Siklósi és Novák (Siklósi és Novák (2015), Siklósi és Novák (2016), Siklósi és Novák (2015)) foglalkoztak a leletek automatikus strukturálásával, szemészeti

ambuláns leletekre fókuszálva, ezen túlmutató munkájuk egy nyelvi modelleken alapuló automatizált javító módszer, és keresőrendszer volt, a strukturálásban a leletek formázását és szózsák modellre alapozó indikátorkifejezéseket felhasználva. Tizenöt, számos esetben szemészet-specifikus címkéjükkel strukturálásuk 81,99%-ban volt helyes. Korábbi munkánk során a radiológiai leletek gépi értelmezésével foglalkoztunk (Kicsi és mtsai (2020)), ennek során a részletes lelet és vélemény szakaszokat használtuk fel entitások detektálására és azonosítására, kapcsolataik feltárlására.

A jelen cikkben bemutatott módszerünk magyar nyelvű klinikai leleteken a mondatok szintjén végez címkézést 5 szakaszba kategorizálással, a szöveg formázásának felhasználása, vagy indikátorkifejezések megállapítása nélkül.

### 3. Módszerek

#### 3.1. Adatok

A feladat során az elsődleges célunk gerincről készült MR-leletek szakaszokra bontása volt. A feladatban egy nagyobb projekt során használt leletekkel dolgoztunk, mely a Budai Egészségközponttal - a továbbiakban [*A intézménnyel*] - közösen kerül megvalósításra. Az ebből a forrásból származó adatokra *A\_teszt*, és *A\_NG* (*nem gerinc-vizsgálat*) néven fogunk hivatkozni. Jelen munkához 600 gerinclelet került véletlenszerű kiválasztásra, melyből 500 leletet használtunk fel a modellek tanításához, és további 100 leletet a különböző adatokon tanult modellek kiértékeléséhez. Rendelkezésünkre álltak továbbá nem gerincről készült leletek is, ezekből 25 került kiválasztásra, melyeket csak a modellek tesztelésénél használtunk fel.

Az annotációs munka több lépésben zajlott. Az első lépés a különböző szakaszok címkéinek kidolgozása volt. Ennek eredményeként öt címke használata mellett döntöttünk, amelyeket bevezetésünk ismertetett bővebben, ezek a következők:

- **Technika:** Minden olyan mondat, mely a vizsgálat típusát (pl. T1 súlyozott, T2, kontrasztanyagos) illetve a vizsgálati régiót írja le.
- **Anamnézis:** Ezen címke alá kerültek a jelenlegi panaszokat, korábbi betegségeket, műtéteket leíró részek, mely azért szükséges, hogy az orvos tágabb kontextusban lássa a páciens problémáját.
- **Részletes lelet:** Ebben a szakaszban kerülnek feljegyzésre az orvosok által megállapított rendellenességek, megállapítások.
- **Melléklet:** Ebbe a csoportba azok a mondatok kerülnek, melyek nem a fő vizsgálati területről tartalmaznak megfigyeléseket.
- **Vélemény:** Tekintható a részletes lelet összefoglalójának, melyben a leletben szereplő fontosabb információk kerülnek megemlítésre.

A tanításhoz használt első 100 lelet kézi felcímkézése után a következő 200 lelet a már meglévő leleteken tanított modellel került előannotálásra, majd az

annotátorok feladata az előannotációk javítása volt. A további annotálandó leletek pedig ezen a 300 leleten tanított modellel kerültek előannotálásra. A leleteket jelen cikk három szerzője annotálta, a kérdéses esetek közös egyeztetésével.

A korábban említett korpusz mellett egy további intézmény, a Szegedi Tudományegyetem - a továbbiakban [*B intézmény*] - klinkáján dolgozó radiológusok által kibocsájtott leletek is rendelkezésünkre álltak. Az ebből a forrásból származó leletekre *B\_teszt* néven fogunk hivatkozni. Tapasztalataink szerint ezek nagymértékben strukturált leletek (legalábbis a szakaszok szintjén), azonban ezek eltérő kifejezésekkel rendelkeznek az [*A intézmény*] adataihoz képest. Ebből a csomagból 25 gerinclelet került felcímkézésre, annak érdekében, hogy más klinikáról származó, eltérő felépítéssel, szókinccsel rendelkező leleteken is le tudjuk mérni a modell pontosságát.

### 3.2. Szabályalapú módszerek

A szakaszok felismerésének feladatára könnyű megoldásnak tűnhetnek a szabályalapú módszerek. A két intézménytől kapott leletek strukturálására egy-egy szabályrendszer került kidolgozásra. Mindkét esetben az első lépés a leletek felépítésének megismerése volt.

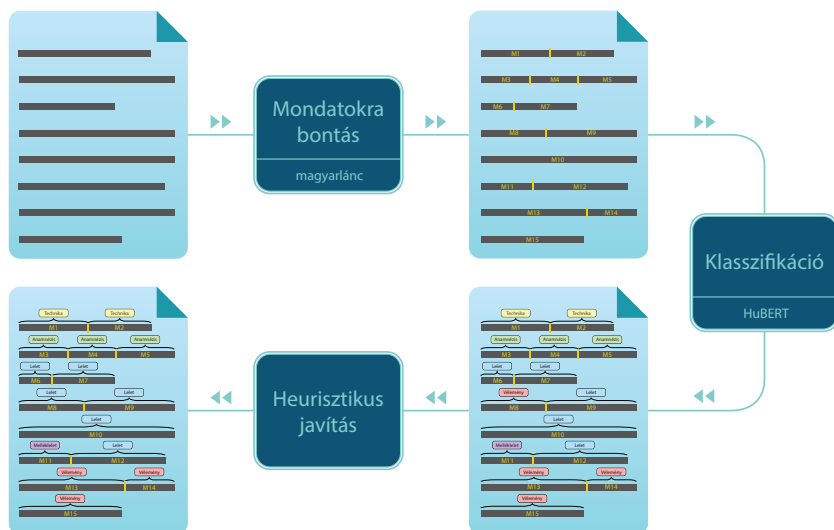
Az [*A intézménytől*] kapott szövegekben gyakoriak voltak az üres sorok, illetve a mondatonkénti újsorok. A leletek sok esetben tartalmaztak bizonyos címkéket a szakaszok elején, mint az anamnézis, vélemény, illetve melléklelet. Ezek elkülönítése könnyű olyan esetekben, amikor például dupla sortörés jelezte a megkezdett címkéhez tartozó rész végét. Azonban azokban az esetekben, amikor az egész lelet nem tartalmazott ilyen formázást, egyszerű szabályokkal próbáltuk megtalálni a szekciók közötti határokat. A címkével nem rendelkező részek meghatározásához a szövegek elemzése során készült szócsoportokat használtuk, melyen belül az egyes csoportok valamely egység gyakori szavait tartalmazták.

A [*B intézménytől*] származó leletek többsége erősen strukturált formával rendelkezett. Itt nem voltak jellemzők a dupla sortörések, azonban szinte minden esetben jelzésre került az új szekció kezdete a szekció nem feltétlenül egységes megnevezésével. A szekcióelnevezések gyűjtése, szinonimák meghatározása után ezen leletek strukturálása már egyszerűen elvégezhető volt a szekciónevek mentén való vágással.

A szabályalapú módszerek fejlesztése általában időigényes, mivel ehhez a szövegek hosszadalmas elemzése szükséges, illetve nehezen alkalmazható más témájú, más intézménytől származó, eltérő felépítésű leletek strukturálásához. A statisztikai módszerek, például szószák modellek segíthetnek ugyan, de rugalmasságuk így is korlátozott.

### 3.3. Hibrid módszer

Módszerünkben a strukturálást mondatonként végeztük, melyben minden mondat egy címke alá sorolható. A kiosztott címkék alapján történik a különböző szekciók meghatározása. Ezt a folyamatot a 2. ábra illusztrálja. A feladat során



2. ábra: Módszerünk illusztrációja

egy hibrid módszer került kidolgozásra, melynek első lépése a leletek megtisztítása és a magyarul (Zsibrita és mtsai (2013)) nyelvi elemzővel való mondatokra bontása. A mondatrabontást követően minden mondatnak elkészítettük a reprezentációját, mely után egy klasszifikációs réteg segítségével kerültek meghatározásra az egyes mondatokhoz tartozó címkék. A modell által létrehozott címkék meghatározását követő lépés a címkék javítása a kiosztott címkék sorrendiségének figyelembevételével, ezt az ábrán heurisztikus javításnak neveztük.

Az egyes mondatokhoz tartozó reprezentációkat a BERT modell felhasználásával készítettük el. A modell egyetlen kimeneti réteg hozzáadásával finomhangolható, így state-of-the-art eredményeket elérve több természetesnyelv-feldolgozási feladaton. Minden, a modell után a folyamatba épített specifikusabb feladathoz (downstream task) külön finomhangolt modell tartozik, így minimális a különbség az előre tanított és a downstream modell architektúrája között. Ma már számos különböző nyelvű és témájú szövegeken előtanított BERT modell létezik, azonban a magyar nyelven tanított BERT modellek választéka kicsi, és az előtanított elérhető modellek általános szövegen kerültek előtanításra. A szabadon használható magyar nyelvű modellek közül a huBERT (Nemeskey (2021)) modell használata mellett döntöttünk. Modellünk architektúrájában a huBERT által készített mondat szintű reprezentációk ([CLS] token) képezték a klasszifikációs réteg bemenetét, mely lehetővé teszi a mondat szintű címkék előállítását.

A következő lépésben a huBERT modellre helyezett klasszifikációs réteggel előállított címkék szabályalapon történő javítására került sor. A szabályokat a kimenetek elemzésével hoztuk létre, és a leggyakoribb hibák javítására fektettük a hangsúlyt. A korábban ismertetett csoportok közül a leghasonlóbb szókinccsel a részletes lelet és a vélemény rendelkezik. A szöveggörnyezet nélkül ezek gyakran

szabad szemmel is megkülönböztethetetlenek. Utófeldolgozási szabályainkban figyelembe vesszük a mondatot körülvevő kiosztott címkéket is. Amennyiben például két azonos címke közé bekerül egy másik típusú címke, a közrevett címkét megváltoztatjuk az őt körülvevőre. Törekedtünk csak olyan minták megalkotására, melyek javítanak a kimeneten, összetett leletek esetében azonban előfordul, hogy helytelenül írjuk át a címkét. Tapasztalataink szerint ezen egyszerű szabályok alkalmazása a pontosság javulását eredményezte. Gyakran előfordult, hogy egy szöveg több leletet is tartalmaz, arról nem is beszélve, hogy más intézményeknél teljesen mások a szokások, ezért a szakaszok sorrendje nem elég jó indikátor.

## 4. Eredmények

Kísérleteink során három adathalmazon történt kiértékelés. Az *A\_teszt* halmaz 100 darab [*A intézménytől*] származó, gerincről készült leletet tartalmaz, az *A\_NG* halmazban 25 darab, szintén az [*A intézménytől*] származó csípő, mellkas vagy térd lelete található, a *B\_teszt* csoportba pedig a [*B intézménytől*] származó gerincleletek kerültek. Kísérleteinkhez pontossági (accuracy) eredményeket közlünk, amelyek a helyesen címkézett mondatok arányát adják meg.

Az 1. táblázat felső szegmensében a tisztán szabályalapú módszerrel elért eredmények láthatók. Az *A\_szabályok* alatt az [*A intézménytől*] kapott leletek elemzése során létrehozott szabályok értendők, míg a *B\_szabályok* a [*B intézménytől*] kapott leletek elemzésével keletkeztek. Ezek is az első bekezdésben említett tesztadatokon lettek kiértékelve.

Az 1. táblázat második szegmensében látható eredmények azt mutatják, hogy milyen hatással volt a tanítóadat mennyiségének bővítése a különböző lelethalmazokon elért pontosságokra, valamint, hogy mekkora mértékű különbség van a csak a modellel előállított, és a heurisztikákkal javított modellkimenetek eredményei között. A táblázatban a *modell\_* kezdetű sorok csak a huBERT segítségével előállított modell, a *hibrid\_* kezdetű sorok pedig a huBERT segítségével előállított modell és az ezen alkalmazott szabályalapú utófeldolgozás eredményeit mutatják. Sorra 100, 300 majd 500 leletből álló tanítóadatot készítettünk. A táblázatból megállapítható, hogy a betanított modellek közül mindhárom teszt-halmaz esetében az 500 leletből álló adaton finomhangolt modellel tudtuk elérni a legnagyobb pontosságot.

A következő lépésben az 500 leletből álló tanítóadatot módosítottuk két különböző módon, az így előállt modellek és a hibrid módszerek kiértékelését pedig a tanítóadathoz hasonlóan módosított, korábban ismertetett teszteken végeztük. Először a leletekből kivágásra kerültek azok a címkéző kifejezések, melyekre a szabályalapú módszerek is építenek, ilyen például az *anamnézis:*, *vélemény:*, *melléklelet:*. Az így előállt tanítóadaton kapott modell a *modell\_500\_1* elnevezést kapta. A másik kísérlet során a korábbi módosítás mellett nem különböztettük meg a részletes lelet és a vélemény címkét, hanem ezeket egy közös címke alá vettük. Ennek magyarázata, hogy általában ezek a valós orvosi megállapítások, a lelet valódi tartalma. A két szakasz mondatai sokszor szabad szemmel



1. táblázat. Különböző adathalmazokon mért pontosság a kialakított szabályalapú módszerrel, a modellel és a hibrid rendszerünkkel

	A_teszt	A_NG	B_teszt
A_szabályok	81,99	79,72	66,85
B_szabályok	0	0,35	93,26
modell_100	95,83	93,30	86,48
hibrid_100	97,74	95,60	87,32
modell_300	96,04	93,65	81,69
hibrid_300	97,81	96,47	81,13
modell_500	97,10	95,06	91,83
hibrid_500	98,23	97,00	92,96

sem különböztethetők meg. Az így készített modell a modell\_500\_2 elnevezést kapta.

A leírtak alapján előállított két korpusz eredményei a 2. táblázatban láthatók. Az 1. táblázat megfelelő soraival összehasonlítva látszik, hogy az így kapott *modell\_500\_1* modell, és az ehhez tartozó hibrid rendszer is alacsonyabb pontosság elérésére képes, míg a másik modell, melyben nem különböztettük meg a leletet és a véleményt az *A\_teszt* halmazon jobb eredményt mutat.

2. táblázat. Az 500 leletből álló tanítóhalmazzal végzett különböző kísérletek

	A_teszt	A_NG
modell_500_1	95,34	90,12
modell_500_1_heur	96,54	90,83
modell_500_2	98,45	95,94
modell_500_2_heur	98,45	95,94

## 5. Az eredmények megvitatása

Bevezetésünkben három kérdés megválaszolását tűztük ki célul. Az első kérdés a mondatonkénti címkézéssel foglalkozott, ennek során megvizsgáljuk a tanítóadat változásának hatásait az eredményekre. Második kérdésünk a nem gerincleteken, és más intézménytől származó leleteken végzett hatást tárgyalja. Harmadik kérdésünk tárgyát a hibrid és a szabályalapú módszerek összehasonlítása képezte.

Munkánk során egy hibrid módszert készítettünk, melynél azt vizsgáltuk, hogy képesek vagyunk-e olyan általános szabályrendszer megalkotására, amely javítani tudja a modell által elért eredményeket. A huBERT modell kifejezetten jól teljesített a feladaton, és az 1. táblázatból az látható, hogy néhány egyszerű

utófeldolgozási szabály alkalmazása szinte minden esetben közel 1%-os javulást eredményezett. Például a *modell\_500* és a *hibrid\_500* sorokat összehasonlítva a hibrid módszer sorra 1,13, 1,94 és 1,13%-os javulást eredményezett. Az eredményekből az is látható, hogy a tanító adathalmaz megfelelő bővítésével jól skálázódik a modell pontossága is. Mivel eltérő tanítóadat esetén a leválasztott 20% tesztadat is eltérne, a modelleket ugyanazon a 100 annotált leleten értékeltük ki, így valóban összehasonlíthatók ezek a számok. Ennek eredményei az 1. táblázat Tesztlelet oszlopához tartozó *modell\_100*, *modell\_300* és *modell\_500* sorokban láthatók. Kísérletünkből kiderül tehát, hogy kevés tanítóadattal készített modellel is jó eredmények érhetők el, melyek tovább javíthatók a modell kimeneteinek elemzése után kialakított szabályokkal.

A 2. táblázat *modell\_500\_1* soraiból az is látható, hogy a leletben lévő, orvosok által beírt címkézést figyelni a huBERT is, de ilyen esetekben is 96% feletti eredményt értünk el. Tehát a címkék figyelése nélkül is sikeresen szétválaszthatók a szakaszok.

**Válasz az 1. kérdésre:** A huBERT-alapú, mondatonként címkéző módszer utófeldolgozással kombinálva, 100 valós gerinccleleten kiértékelve 98,23-as, a leletekben lévő címkék nélkül 96,54-es pontosságot produkált.

Második kérdésünk központjában az állt, hogy más adatokon is képes-e jól teljesíteni a módszer, melynek eredményei az 1. táblázat sorainak *A\_NG* és *B\_teszt* oszlopaiból olvashatók le. Az itt látható eredményeken azt láthatjuk, hogy módszerünkkel nem csak az [*A intézmény*] gerinccleletein, hanem más régiókról készült leleteken, és más intézmény leletein is képes volt a strukturálásra. Érdekes, hogy a 300 leletes tanítás esetében a [*B intézmény*] leletein jelentős romlás látszik a kisebb tanítóhalmazhoz képest. Ez vizsgálatunk alapján valószínűleg annak tudható be, hogy a tanítóhalmaz random bővítésével csökkentettük az anamnézist tartalmazó címkék arányát a többi címkéhez képest, mely negatív hatással volt a [*B intézmény*] leleteire. Az 500 leletes tanításhoz a 300 leletet olyan leletekkel bővítettük, melyek tartalmazznak anamnézis és melléklelet szekciókat is, így javítva a [*B intézmény*] leleteinek strukturálását.

**Válasz a 2. kérdésre:** Módszerünk egy intézmény gerinccletein finomhangolva más anatómiai régiókon (97,00), és más intézmény gerinccletein (92,96) is nagy pontossággal működik, láthatóan jól általánosítva.

A harmadik kérdésre, a pusztán szabályalapú módszer eredményeit a modell és a hibrid eljárás eredményeivel összehasonlítva kapjuk meg a választ. Ehhez az 1. táblázat sorainak vizsgálata szükséges. Megfigyelhető, hogy az [*A intézménytől*] származó gerinccleletekre kialakított szabályalapú (*A\_sabályok*) módszer eredményei minden esetben elmaradnak még a legkevesebb adaton készített hibrid rendszer eredményeihez képest is. A [*B intézmény*] leleteinél az intézményhez kialakított szabályalapú módszert a modellnek nem sikerült felülmúlnia, de nagyon kis különbséggel megközelítette. Az is látható, hogy annak ellenére, hogy ebben az adathalmazban jobban strukturált leletek vannak, a 100%-os eredmény szabályokkal is elmarad, mivel rosszul címkézett, vagy címkézetlen leletek esetén a kialakított szabályok nem alkalmazhatók. A 0, és közel 0%-os eredményből lát-

szik, hogy más intézmény leletein, mely más címkéket tartalmaz, vagy teljesen strukturálatlan leletek esetén ez a szabályrendszer alkalmazhatatlan.

**Válasz a 3. kérdésre:** Kísérleteink alapján az általános módszer megközelíti, vagy túl is szárnyalja az intézményekhez optimalizált, meglévő címkékre alapozó szabályokat.

Az utófeldolgozás, azaz a modell kimenetét a környezet alapján korrigáló szabályok sokkal egyszerűbben és gyorsabban létrehozhatók, mintha teljesen szabályalapon szeretnénk a leletek strukturálását végezni. Emellett egy modelltől és szabályokból álló hibrid módszer jobb általánosításra képes, amennyiben más területet leíró illetve más felépítésű leletek strukturálására is szükség van. Az utófeldolgozás (hibrid módszer) az összes mérésnél javítást eredményezett, kivéve a modell\_500\_2 esetében, ahol az eredmény változatlan. Ez a kivétel azzal magyarázható, hogy a legtöbb jól javítható hiba a lelet és vélemény elkülönítésében volt, mivel ezek valóban nagyon hasonló mondatok, és ennél a halmagnál ezt a két címkét nem különböztettük meg. Megjegyezzük, hogy ez is a valós használatnak is megfelelő eset lehet, mert ezek a leletezés legfontosabb kimenetei, a módszer jól szegmentálja ezeket a szakaszokat a többitől.

## 6. Összegzés

Cikkünkben magyar nyelvű gerincelemek automatikus strukturálásával foglalkoztunk, a leletet mondatonkénti címkézéssel szakaszokra bontva. Kidolgoztunk két szabályalapú módszert intézményekre optimalizálva, valamint egy huBERT reprezentációt felhasználó modellt, és utófeldolgozással hibrid módszert készítettünk. Az eredmények azt mutatják, hogy a huBERT-alapú klasszifikációs modell 98,23% pontossággal állapítja meg a szakaszokat, jó eredményeket adva akkor is, ha az adat nem is tartalmaz címkéket az orvosoktól (96,54% ebben az esetben): A módszer jól általánosít más intézményekre (92,96%) vagy anatómiai régiókra (97.00%), illetve megközelíti, vagy felül is múlja az intézményekre optimalizált szabályalapú módszereket. A jelenlegi megoldás több fejlesztési lehetőséget kínál, mint például a tanítóadat más intézményekből származó leletekkel való bővítése, illetve a hibrid rendszer szabályainak további pontosítása. Munkánkban több felhasználása lehet, alkalmas például más kutatási/adatbányászati alkalmazások készítéséhez szükséges adatok tisztítására, vagy a leletek jobb rendszerezésére intézményeken belül.

## Köszönetnyilvánítás

A kutatás az Európai Unió támogatásával valósult meg, az RRF-2.3.1-21-2022-00004 azonosítójú, Mesterséges Intelligencia Nemzeti Laboratórium projekt keretében. A kutatás továbbá a Kulturális és Innovációs Minisztérium ÚNKP-22-4-SZTE-500 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.

## Hivatkozások

- Apostolova, E., Channin, D., Demner-Fushman, D., Furst, J., Lytinen, S., Raicu, D.: Automatic segmentation of clinical texts. In: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. vol. 2009, pp. 5905–5908 (09 2009)
- Bosmans, J.M.L., Weyler, J.J., De Schepper, A.M., Parizel, P.M.: The radiology report as seen by radiologists and referring clinicians: Results of the cover and rover surveys. *Radiology* 259(1), 184–195 (Apr 2011)
- Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bi-directional transformers for language understanding. In: NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. p. 4171–4186 (2019)
- Ganeshan, D., Duong, P.A.T., Probyn, L., Lenchik, L., McArthur, T.A., Retrovey, M., Ghobadi, E.H., Desouches, S.L., Pastel, D., Francis, I.R.: Structured reporting in radiology. *Academic Radiology* 25(1), 66–73 (Jan 2018)
- Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification (Aug 2016), arXiv preprint, arXiv:1607.01759
- Kicsi, A., Pusztai, P., Szabó Ledenyi, K., Szabó, E., Berend, G., Vincze, V., Vidács, L.: Információkinyerés magyar nyelvű gerinc MR leletekből. In: XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019). p. 177–186. Szeged (2019)
- Kicsi, A., Szabó Ledenyi, K., Pusztai, P., Németh, P., Vidács, L.: Entitások azonosítása és szemantikai kapcsolatok feltárása radiológiai leletekben. In: XVI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2020). p. 15–28. Szeged (2020)
- Kuling, G., Curpen, B., Martel, A.L.: Bi-rads bert and using section segmentation to understand radiology reports. *Journal of Imaging* 8(5), 131 (May 2022)
- Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space (Sep 2013), arXiv preprint, arXiv:1301.3781
- Nemeskey, D.M.: Introducing huBERT. In: XVII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2021). p. 3–14. Szeged (2021)
- Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Empirical Methods in Natural Language Processing (EMNLP). pp. 1532–1543 (2014)
- Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations (Mar 2018), arXiv preprint, arXiv:1802.05365
- Pomares-Quimbaya, A., Kreuzthaler, M., Schulz, S.: Current approaches to identify sections within clinical narratives from electronic health records: a systematic review. *BMC Medical Research Methodology* 19(1), 155 (Jul 2019)
- Rosenthal, S., Barker, K., Liang, Z.: Leveraging medical literature for section prediction in electronic health records. In: 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). p. 4863–4872. Association for Computational Linguistics, Hong Kong, China (2019)

- Sadoughi, N., Finley, G.P., Edwards, E., Robinson, A., Korenevsky, M., Brenndorfer, M., Axtmann, N., Miller, M., Suendermann-Oeft, D.: Detecting section boundaries in medical dictations: Toward real-time conversion of medical dictations to clinical reports. In: International Conference on Speech and Computer. p. 563–573. Springer International Publishing, Cham (2018)
- Siklósi, B., Novák, A.: Rec. et exp. aut. Abbr. nyelv. KLIN. szövegben – Rövidítések Automatikus Felismerése és Feloldása Magyar Nyelvű Klinikai Szövegekben. In: X. Magyar Számítógépes Nyelvészeti Konferencia. p. 167–176. Szeged (2014)
- Siklósi, B., Novák, A.: Nem Felügyelt Módszerek Alkalmazása Releváns Kifejezések Azonosítására és Csoportosítására Klinikai Dokumentumokban. In: XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015). p. 237–248. Szeged (2015)
- Siklósi, B., Novák, A.: Digitális Konzílium – Egy Szemészeti Klinikai Keresőrendszer. In: XII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2016). p. 230–240. Szeged (2016)
- Siklósi, B., Novák, A.: Restoring the intended structure of hungarian ophthalmology documents. In: Proceedings of BioNLP 15. pp. 152–157. Association for Computational Linguistics, Beijing, China (July 2015)
- Taira, R.K., Soderland, S.G., Jakobovits, R.M.: Automatic structuring of radiology free-text reports. *RadioGraphics* 21(1), 237–45 (Jan 2001)
- Tepper, M., Capurro, D., Xia, F., Vanderwende, L., Yetisgen-Yildiz, M.: Statistical section segmentation in free-text clinical records p. 2001–2008 (2012)
- Zsibrita, J., Vincze, V., Farkas, R.: magyarlanc: A toolkit for morphological and dependency parsing of hungarian. In: Proceedings of RANLP. p. 763–771 (2013)