

# Digitális Konzílium – egy szemészeti klinikai keresőrendszer

Siklósi Borbála<sup>1</sup>, Novák Attila<sup>1,2</sup>

<sup>1</sup> Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar,

<sup>2</sup> MTA-PPKE Magyar Nyelvtechnológiai Kutatócsoport

1083 Budapest, Práter utca 50/a

e-mail: {siklosi.borbala,novak.attila}@itk.ppke.hu

**Kivonat** A klinikai dokumentumok feldolgozása a nyelvtechnológia egyik kiemelkedő és igen hasznos alkalmazási területe. A klinikai körülmények között létrejövő beteglapok igen sok hasznos információt tartalmaznak a beteg mellett az orvosok számára is. Ezek tárolási módja azonban nem teszi lehetővé ezeknek az információknak az elérését.

Cikkünkben egy magyar nyelvű szemészeti dokumentumokat feldolgozó láncot és a feldolgozott dokumentumokra épülő összetett keresőrendszer első változatát mutatjuk be. Az alkalmazott módszerek a klinikai dokumentumok sajátosságait veszik figyelembe az előfeldolgozás első lépéseitől kezdve a keresőfelület kialakításáig.

## 1. Bevezetés

A klinikai dokumentumok, beteglapok feldolgozása a nyelvtechnológia egyik fontos alkalmazása, amely egyre jobban elválnak az orvosbiológiai szövegekre épülő alkalmazásoktól [4,10,3]. A kórházi körülmények között létrejövő dokumentumok sok hasznos információt tartalmazhatnak más orvosok vagy szakértők számára. Sokszor azonban a dokumentumok formátuma és tárolási módja, illetve a létrehozásukat „támogató” eszközök ellehetetlenítik a tényleges tartalomhoz való hozzáférést, elzárva ezzel az orvosok előtt ezt a gazdag információforrást.

A klinikai dokumentumok létrehozásának módszere két fő csoportba sorolható. Az egyik esetben egy EHR (Electronic Health Records) rendszeren keresztül kerülnek rögzítésre. Ekkor az orvosok vagy az asszisztensek egy, a rendszer által felajánlott sablont töltenek ki, melynek eredménye valamilyen strukturált dokumentum. A dokumentum szerkezetének részletessége függ a dokumentációs rendszertől és a felhasználói szokásoktól is. A klinikai dokumentumok létrehozásának másik módja a hagyományos, kézzel írt dokumentációs szokásokhoz hasonlítható. Ekkor, bár a dokumentumok létrehozása és tárolása is számítógépen történik, ez tulajdonképpen csak mint egy írógép használható. A létrejött dokumentum pedig csupán egy egyszerű szövegfile, amiben a dokumentum tartalmi szerkezetének nyomai csak a szintén kézzel rögzített formázási jegyekben fedezhetők fel.

Cikkünkben egy olyan rendszer prototípusát mutatjuk be, ami magyar nyelvű szemészeti dokumentumok egy gyűjteményét alakítja át kereshető, strukturált

formába, és az így feldolgozott leletanyaghoz keresőfelületet is biztosít. A magyar kórházakban az EHR rendszerek használata hagy némi kívánnivalót maga után. Bár a legtöbb esetben használnak ilyen rendszereket, az ezeket használó orvosok és asszisztensek a rendszer rugalmatlansága és bonyolultsága miatt inkább saját dokumentációs szokásaikat tartják meg, a rögzített információt a felkínált sablon egyetlen mezőjében rögzítve.

Különösen nehéz terület a szemészeti dokumentáció, ahol az EHR rendszerek alkalmazása más országokban sem jár igazán sikerrel [1,9,2]. A sokféle mérési eredmény rögzítésének módja (amik közül néhány táblázatos formát igényel, a többbit egyetlen szám vagy rövid szöveg, vagy akár ábra ír le) olyan rendszert igényelne, aminek a létrehozása és használata igen nehéz.

A szemészeti dokumentáció egy másik sajátos tulajdonsága, hogy a feljegyzések sietve, a vizsgálat közben jönnek létre. Ezért igen jellemzőek a gyakran ad hoc rövidítések, az elírások, az angol, a latin és a magyar nyelv vegyes használata, illetve a szöveges leírásokban is gyakran hiányos nyelvtani szerkezetek jönnek létre. Ezért a legtöbb, általános magyar szövegekre jól alkalmazható elemzőrendszer nem használható közvetlenül ezeknek a szövegeknek az elemzésére. Ez a feldolgozási lánc olyan alapvető elemeire is igaz, mint a tokenizálás, mondathatárfelismerés, vagy szófaji egyértelműsítés. Az 1. ábra egy angol, míg a 2. ábra egy magyar nyelvű szemészeti leletet mutat be, amikben jól látszanak a szemészeti dokumentáció során használt nyelvezet sajátosságai. A leletek feldolgozása tehát nem csak a magyar nyelvből fakadó összetettség miatt nehéz, hanem a szemészeti domén jellemzője bármely nyelv esetén.

```

va wc
od 20/ 60 ph 20/50-2
os 20-/100 ph ni stable
p 4-2 reactive ou no rapd
eom full
sle: lla mild blepharitis ou
c.s 1+ papillary rxn ou
k inf spk ou
ac d+q ou
i rr ou no rubeosis
l 2+ ns ou brunescant
ta 14 ou
pp m1/m2.5
c:d 0.3 ou
no bdr

```

1. ábra. Egy angol nyelvű szemészeti lelet részlete

A bemutatott rendszer célja, hogy mindezeket a nehézségeket kezelni tudja. A feldolgozás során a nyers szöveges dokumentumokból indul ki, amikben először a helyesírási hibák automatikus javítása történik, majd a rövidítések felismerése és feloldása, és a kórtörténet, illetve a dokumentumstruktúra automatikus meghatározása, melynek során minden dokumentum minden sora egy adott tartalmi

kategóriába sorol be a rendszer. Végül az így feldolgozott és strukturált dokumentumokat indexeljük, és egy speciális keresőfelületen keresztül kereshetőek.

## 2. A magyar szemészeti korpusz

Vizsgálataink során szemészeti osztályon keletkezett anonimizált magyar nyelvű dokumentumokat használtunk. A dokumentációs rendszer hiánya, vagy annak elnagyolt használata miatt a dokumentumok struktúrájára az esetek nagy részében csak az esetleges (a korpuszban nem is egységesen használt) formázási jegyek, illetve a dokumentumok tartalmának értelmezése utal. Továbbá, a feljegyzések jelentős részének tárolása redundáns módon történt, a teljes vagy részleges kórtörténet többszöri, kézzel történő másolása miatt. A dokumentumok szöveges része nagyon sok helyesírási hibát, rövidítést, idegen nyelvű terminológiát tartalmaz, amiknek a lejegyzése nem követ semmilyen helyesírási normát. A 2. ábrán egy példadokumentum látható az eredeti formájában.

A M B U L Á N S   K E Z E L Ő L A P

Státusz  
2010.10.19 12:28

Olvasó szemüveget szeretne. Néha könnyeznek a szemei.  
V:0,7+0,75Dsph=1,0  
1,0 +0,5 Dsph élesebb

+2.0 Dsph mko Cs IV

St. o. u: halvány kh, ép cornea, csarnok kp mély tiszta, iris ép békés, pupilla rekciónk rendben, lencse tiszta, jó vvf.  
Átfecskendezés mko sikerült.

Olvasó szemüveg javasolt: +2.0 Dsph mko.  
Éjszakánként műkönyggél ha szükséges.  
Kontroll: panasz esetén

Diaagnózis	Kód	Dátum	Év	K	V	T
DIAGNÓZISOK megnevezése	H5390	2010.10.19				
Látászavar, k. m. n.				3		

Beavatkozások	Menny.	Pont
Kód Megnevezés		
11041 Vízgálat	1	750

2010.11.16

2. ábra. Egy szemészeti lelet annak eredeti formájában

A szemészeti dokumentumok egyedi sajátossága a szöveges részek közé illesztett táblázatos, vagy áltáblázatos adat. Ezek a szövegfeldolgozás során zajt képeznek az adott környezetben. Ilyen nem szöveges információk a laboreredmények, a szemvizsgálat során mért látásélességi mérések eredményei, illetve speciális elválasztó karakterek, vagy más speciális karakterek sorozatai. A dokumentumok ezen részeire is jellemző a szabványos forma hiánya, így dokumentumról dokumentumra (vagy orvosról orvosra) változó lehet a leírás módja.

A dokumentumok szöveges részeit tekintve is számos jelentős különbség fedezhető fel az általános magyar nyelvű szövegek és a szemészeti domén között,

ami magyarázatot ad arra, hogy az általános szövegeken akár bonyolultabb feladatok esetén is jól teljesítő eszközök miért nem alkalmazhatóak a klinikai dokumentumokra. A különbség a két szövegtípus között nem csak azok tartalmában nyilvánul meg, hanem már a nyelvtani szerkezetekben és a szövegekben előforduló szóalakokban is. A két domén részletes összehasonlítása megtalálható [12]-ben és [14]-ben.

### 3. Előfeldolgozási lépések

A fent ismertetett sajátos, igen zajos és a bármilyen helyesírási és szerkesztési normához való igazodást nélkülöző nyelvezettel lejegyzett klinikai dokumentumok automatikus feldolgozása igen nehéz, ezért az előfeldolgozási lánc létrehozása és adaptálása minden további lépés előfeltétele és komoly munkát igényel. Az összes előfeldolgozási lépés részletes ismertetése meghaladja a jelen cikk korlátait, ezért csak röviden ismertetjük őket.

A **mondatok szegmentálása és a tokenizálás** az első alapvető lépés a szöveges részek feldolgozása során. Ennek megvalósítása két lépésben történik, először a tokenizálás, aztán pedig a mondatvéget jelző indikátorok log-likelihood alapú osztályozásával a mondathatárok felismerése. A módszer eredményessége 96,0%-os F-mértékkel jellemezhető a tokenizálásra és 91,89%-os F-mértékkel a mondathatár-felismerésre. [6]

A **helyesírási hibák automatikus javítása** során célunk egy kváziszten-derd forma elérése volt a szöveges részekben. Ehhez egy statisztikai gépi fordítón alapuló módszert alkalmaztunk, de a rendszer párhuzamos korpuszból való tanítása helyett a fordítási modellt egy javaslatgeneráló rendszerrel helyettesítettük, ami minden szóalakhhoz lehetséges javításjelölteket generált. A dekóder feladata ezek közül a megfelelő alak kiválasztása, a nyelvmodell pedig a szöveggörnyezetet képviselve súlyozza a lehetséges alakok valószínűségét. A módszer 87,23%-os pontossággal működik.[14]

Már a tokenizálás és szegmentálás feladatát is jelentősen megnehezítette a rengeteg rövidítés, illetve azok igen sokféle megjelenési formája. Egy kezelhető reprezentáció elérése érdekében pedig különösen fontos lépés a jelentéssel bíró egységek azonosítása, így a **rövidítések felismerése és feloldása**. [15] A szövegek normalizálása során történő azonosítás és feloldás mellett a felderített rövidítéseket és feloldásait később a keresőrendszerbe is integráltuk, szinonimahalmazokként megadva őket, ezzel lehetővé téve a rövidített és kifejtett előfordulások megfeleltetését a keresés során.

Az előfeldolgozási lánc utolsó eleme a dokumentum szöveges részeinek **szó-faji egyértelműsítése**, melyre a PurePos adaptált verzióját használtuk [7]. Az egyértelműsítő teljesítménye ugyan elmarad az általános szövegeken mérhetőtől, de a klinikai szövegeken is 90% fölötti pontosságot ért el.

A fenti lépések alkalmazása után létrejött egy kvázi-normalizált, illetve lemmatizált és morfoszintaktikai címkékkel ellátott reprezentáció a dokumentumok szöveges részeire. Ezután a rendszerhez **szemantikai információkat** adtunk

hozzá. Mivel magyar nyelvre nem nagyon létezik használható a szemészeti domént megfelelően reprezentáló lexikai szemantikai erőforrás, ezért egy olyan módszert alkalmaztunk, amivel a korpuszból létrehozott disztribúciós szemantikai modell alapján definiáltunk egy szemantikailag releváns fogalmi rendszert [11].

## 4. Dokumentumszerkezet azonosítása

A dokumentum szerkezetének azonosítását két lépésben valósítottuk meg. Először egy a formázási jegyeket azonosító algoritmust alkalmaztunk, majd a dokumentumban szereplő minden sort egy-egy tartalmi egységet jelölő címkével azonosított osztályba soroltunk be [13]. Az 1. táblázat az osztályozáshoz definiált tartalmi egységeket foglalja össze.

1. táblázat. A dokumentum állításainak kategorizálásához használt kategóriacímkék és jelentésük

címke	jelentés	leírás
Tens	Szemnyomás	A szemnyomás mérésének eredménye
V/Refr	Refrakció	Refrakció adatok
Ana	Anamnézis	A beteg panaszai, korábbi betegségek, családi kórtörténet, stb.
Dg	Diagnózis	Az aktuális diagnózis
Beav	Beavatkozás	Alkalmazott kezelés, kivéve ha műtét vagy gyógyszeres terápia
Vél	Vélemény	Az orvos által megfogalmazott vélemény, kivéve ha diagnózis vagy kezelés
St	Státusz	A beteg aktuális állapota
Ther	Terápia	Felírt vagy alkalmazott gyógyszeres kezelés
BNO	BNO	Betegségekhez vagy kezelésekhöz tartozó BNO kódok
T	Teszt	Alkalmazott vizsgálatok, kivéve a R1 kategóriába tartozók
V	Visus	Látásélesség vizsgálatának eredményei
R1	Réslámpa	Réslámpával végzett vizsgálatok
Kontr	Kontroll	Kontrollvizsgálatra vonatkozó információk
Műtét	Műtét	Előírt vagy végrehajtott műtétek
XXX	-	A fenti kategóriák egyikébe sem sorolható állítások

### 4.1. Kórtörténet azonosítása

Igen gyakran előfordul, hogy az aktuális beteglap egyes részei a betegre vonatkozó korábbi dokumentumokból származó részleges vagy teljes másolatok, esetenként kisebb változtatásokkal. Így bár ezek a részletek csupán redundáns információt tartalmaznak (a teljes betegtörténetre vonatkozóan), felismerésükre nem volt elegendő a teljes egyezés vizsgálata. A másik nehézséget ezek azonosítása során az jelentette, hogy sok olyan feljegyzésrészlet is van, ami közel minden

vizsgálatnak része, és gyakran tartalmilag is hasonlóak, így egy egyszerű illeszkedésvizsgálat során tévesen azonosíthatnánk másolatként ezeket. A tényleges másolatok és változataik felismerésére ezért egy olyan módszert alkalmaztunk, melynek alapja a dokumentumok adott részeinek md5 kódolt változatának összehasonlítása [13]. A módszer alkalmazása után az egyes betegekhez tartozó kórtörténet visszakövethetővé vált, az eredetileg egyetlen dokumentumban összefűzve tárolt információ időrendi sorrendben kinyerhető és egységekre bontható.

## 4.2. Feljegyzett sorok osztályozása

A formázási jegyek alapján azonosított nagyobb szerkezeti egységeken belül is többféle tartalmi egységbe sorolható információ került feljegyzésre a dokumentumokban. Ezek azonosítását két lépésben oldottuk meg.

Először a szövegek előfeldolgozott változata alapján definiáltunk a szófajcímkék és a leggyakoribb kifejezésekhez rendelt szemantikai kategóriacímkék alapján olyan mintázatokat, amik az adott állítást valamilyen tartalmi egységbe sorolták. Például az igék alapvetően ritka használata miatt, ha egy múlt idejű igealak környezetében bizonyos szemantikai csoportba tartozó szavak kerültek, akkor az adott mondat nagy valószínűséggel az anamnézis kategóriába sorolható.

A második lépésben a többi sort osztályoztuk. Ehhez először az egyes tartalmi egységekhez indikátorszavakat gyűjtöttünk a korpuszból (néhány példa látható a 2. táblázatban). Az adott kategóriához tartozó indikátorszót tartalmazó sorokat megcímkéztük a kategóriacímkével, majd az így azonosított sorokból minden kategóriához egy szózsák modellt hoztunk létre tf-idf súlyozással. A címkézetlenül maradt sorokat pedig ezekhez a modellekhez hasonlítva kategorizáltuk [13].

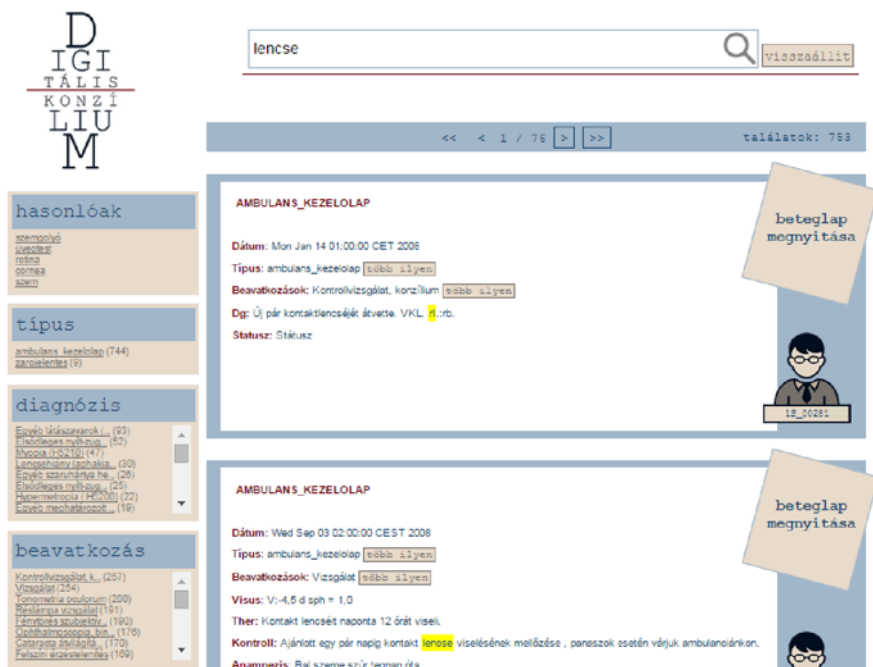
2. táblázat. Néhány példa a címkékre és a hozzájuk tartozó indikátorkifejezésekre

Ana	T	RL	Ther
egyéb betegség	eredmény	réslámpa	th
család	ultrahang	macula	szemcsepp
korábbi	Topo	fundus	terápia
hypertonia	Schirmer	rl	rendelés
anamnézis		lencse	javasolt
panasz			

## 5. A keresőrendszer

Célunk egy olyan keresőrendszer megvalósítása volt, amiben a dokumentumok strukturált és feldolgozott állapotban kerülnek eltárolásra és válnak kereshetővé oly módon, hogy akár tartalmi egységre, akár más szempont alapján való megszorítások, illetve összetett lekérdezések is megfogalmazhatóak legyenek. Ezért

a tartalmi címkékkel megjelölt dokumentumokat egy olyan XML-sémának megfelelően alakítottuk át, ami az egyes tartalmi egységeket elkülöníti, de a dokumentumok eredeti formáját is megőrzi. Az összes dokumentum így átalakított formáját egy Solr alapú keresőrendszerben indexeltük és kereshetővé tettük. A tárolt szerkezethez, illetve a dokumentumok sajátosságaihoz és a keresőrendszer eredeti célkitűzéseivel illeszkedő keresőfelület a 3. ábrán látható.



3. ábra. A keresőrendszerhez készült keresőfelület prototípusa

## 5.1. Csoportosítás

Az összetett, ugyanakkor rugalmas és könnyen kezelhető keresőfelület megvalósítása érdekében a rendszer különböző szempontok szerinti dinamikus csoportosítást valósít meg az aktuális keresés során kapott eredménylista alapján. Így a találatok halmaza tovább szűkíthető az azokban található diagnózisok, vizsgálatok és alkalmazott kezelések mentén. Továbbá, szintén a keresés lefuttatása után a felületen állítható a találati eredmények időbeli korlátozása is.

A találati lista szűrése mellett pedig kapcsolódó kifejezéseket is hozzáadhatunk az eredeti lekérdezéshez. Ezekre a rendszer tesz ajánlatot, miután a felhasználó beírta az általa keresett kifejezést. Az ajánlatok listája a szemészeti korpuszból korábban létrehozott fogalmi rendszer alapján jön létre, melyek akár

többszavas kifejezéseket vagy rövidített alakokat is tartalmazhat. A kategorizáció a fogalmak disztribúciós modellje alapján hierarchikus klaszterezéssel jött létre, melynek során a teljes hierarchiából koherens, tömör csoportok kivágásával jöttek létre az egymáshoz kapcsolódó kifejezések halmazai [16]. Így tehát a keresőkifejezés beírása után az azzal egy csoportban szereplő kifejezéseket ajánlja fel a rendszer, melyek nem csupán szinonimák, hanem lazábban kapcsolódó kifejezések is lehetnek. Például a *lencse*, *szemgolyó*, *üvegtesti tér*, *retina*, *kornea*, *szem* egy csoportba került kifejezések, a keresés során pedig jól finomíthatják vagy kiegészíthetik egymást, az automatikus felajánlásuk pedig megkönnyíti a felhasználó dolgát.

## 5.2. Szinonimák és rövidítések

A szinonimák azonosítása mind az indexelés, mind a lekérdezés során megtörténik. A rendszerben tárolt szinonimák meghatározása szintén a korpusz alapján történt. Szinonimaként tároljuk továbbá a rövidítések különböző változatait és azok feloldásait is, melyeket az előfeldolgozás során határoztunk meg. Például a *mindkét szem* kifejezés lekérdezése során a találati listában ennek a kifejezésnek az összes változatára illeszkedő dokumentumokat találunk. Így a magyar rövidített alaknak megfelelő *mksz* (és változatai), vagy a latin alaknak megfelelő rövidített változatok *ou*, *o.u.*, *o.utr.* *OU* (oculi utriusque ‘mindkét szem’) is illeszkedő találatok lesznek.

A rövidítések szinonimaként való kezelése azonban zajt is eredményezhet a rendszerben, a nagyon sokféle és nagyon gyakori rövidítések miatt. Ezért nem egyetlen általános szinonimalistát tárolunk, hanem a tartalmi egységeknek megfelelően a rövidítéseket is osztályoztuk. Így ha egy rövidítés az egyik tartalmi egységben gyakori, míg a többiben nem használatos, akkor azt csak az ahhoz az egységhez tárolt és használt szinonimalistában tároljuk.

## 5.3. Helyesírási változatok

Mivel a rendszert magyar nyelvű dokumentumok keresésére alkalmazzuk, a helyesírási változatok kezelése nem egyszerű, de igen fontos feladat. Bár a dokumentumok előfeldolgozása során alkalmaztunk egy modult a helyesírási hibák javítására, a lekérdezés ellenőrzése csak online történhet. Ennek megoldására a Humor [8,5] morfológiai elemző orvosi terminológiával kiegészített változatát [7] integráltuk a rendszerbe.

## 5.4. Kórtörténet és beteglapok megjelenítése

A találati lista megjelenítésekor a dokumentumok feldolgozott és strukturált változata jelenik meg, amiben minden sor valamilyen tartalmi egységbe került besorolásra. A keresőkifejezésre való illeszkedést pedig kiemeléssel jelöljük. A felhasználónak azonban lehetősége van az eredeti dokumentum megjelenítésére is, annak eredeti alakjában. Mivel a rendszer megcélzott felhasználói köre az orvosok, ezért különösen fontos, hogy a számukra ismert és használt megjelenítést



is biztosítsunk. Amikor a felületen erre a megjelenítési módra váltunk, akkor az aktuális dokumentum egy léptethető időskálán jelenik meg, melyen így a kórtörténet átláthatóvá és visszakövethetővé válik közvetlenül a felületről. Az időskála bármely pontjáról visszatérhetünk az eredeti találati listához.

### 5.5. A *visus* mező

Mivel a keresőrendszer célzottan a szemészeti kezelőlapokhoz készült, ezek egyik jellemző sajátossága a beteg szemvizsgálata során kapott optikai jellemzők táblázatszerű feljegyzése (*visus*). Ennek során a beteg látását javító dioptria-, látásélesség-, szemnyomás- stb. értékeket az orvosok saját szokásaiknak megfelelő, általában semmilyen szabványhoz nem igazodó formában írják le. Sok esetben azonban ezek azok a részek az egyes dokumentumokban, amik a legfontosabb információkat tartalmazzák, a diagnózis is sok esetben nagyrészt ezekből határozható meg. Mivel ezek az adatok különböző, általában diszkrét skálákhoz igazodnak, ezért a keresés során a konkrét értékek mellett azok tartománybeli elhelyezkedése is fontos lehet. Mintaillesztési módszerek alkalmazásával megvalósítottunk egy olyan funkciót, amelyben akár szövegesen, akár konkrét értékekkel megadott keresés illeszthető a dokumentumokban különböző formátumban tárolt numerikus *visus*értékekre. Például a *közepes fokú rövidlátás* keresőkifejezés találatai között megtaláljuk azokat a dokumentumokat, ahol a betegnél mért szferikus dioptriaérték negatív előjelű, és a megfelelő tartományba esik. Ezzel egy időben a keresőrendszer többi funkciói is aktívak, így pl. az előbbi kifejezés valamely rövidített latin megfelelőjét (pl. *myop.med.grad.*) tartalmazó dokumentumok is találatok maradnak.

## 6. Konklúzió

Cikkünkben egy olyan keresőrendszer első verzióját mutattuk be, amely szemészeti dokumentumokhoz biztosít intelligens keresési felületet. A szemészeti kórlapok a klinikai dokumentumokon belül is speciálisan nehezen kezelhető részhalmozat jelentenek, ezért több előfeldolgozási lépés kifejlesztésére és alkalmazására volt szükség. Az előfeldolgozott dokumentumokat és az azokhoz tartozó hozzáadott információkat (rövidítések, kapcsolódó kifejezések, írásváltozatok stb.) egy keresőrendszer számára indexelhetővé tettük. Bemutattuk továbbá azt a keresőfelületet, ahol a dokumentumokban való keresésre összetett lekérdezéseket adhatunk meg, illetve a találati listát szűkíthetjük, bővíthetjük, a dokumentumok pedig több nézetben is megjeleníthetők.

Jelen cikkünkben a keresőrendszer első változatát mutattuk be, amely még természetesen további fejlesztésre szorul, de már ebben az állapotában is hiánypótló eszköz lehet a szemészettel foglalkozó orvosok számára, amit gyakorló szemészorvosok is megerősítettek.

## Hivatkozások

1. Chiang, M.F., Read-Brown, S., Tu, D.C., Choi, D., Sanders, D.S., Hwang, T.S., Bailey, S., Karr, D.J., Cottle, E., Morrison, J.C., Wilson, D.J., Yackel, T.R.: Evaluation of electronic health record implementation in ophthalmology at an academic medical center (an american ophthalmological society thesis). *Trans Am Ophthalmol Soc* 111, 70–92 (Sep 2013)
2. Elliott, A., Davidson, A., Lum, F., Chiang, M., Saaddine, J.B., Zhang, X., Crews, J.E., Chou, C.F.: Use of electronic health records and administrative data for public health surveillance of eye health and vision-related conditions. *Am J Ophthalmol* 154(6 0), S63–S70 (Dec 2012), <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4037529/>
3. Friedman, C., Johnson, S., Forman, B., Starren, J.: Architectural requirements for a multipurpose natural language processor in the clinical environment. *Proc Annu Symp Comput Appl Med Care* pp. 347–51 (1995)
4. Meystre, S., Savova, G., Kipper-Schuler, K., Hurdle, J.: Extracting information from textual documents in the electronic health record: a review of recent research. *Yearb Med Inform* 35, 128–44 (2008)
5. Novák, A.: Milyen a jó Humor? In: I. Magyar Számítógépes Nyelvészeti Konferencia. pp. 138–144. SZTE, Szeged (2003)
6. Orosz, Gy., Novák, A., Prószéky, G.: Hybrid text segmentation for Hungarian clinical records, *Lecture Notes in Artificial Intelligence*, vol. 8265. Springer-Verlag, Heidelberg (2013)
7. Orosz, Gy., Novák, A., Prószéky, G.: Lessons learned from tagging clinical Hungarian. *International Journal of Computational Linguistics and Applications* 5(1), 159–176 (2014), <http://www.gelbukh.com/ijcla/2014-1/>
8. Prószéky, G., Kis, B.: A unification-based approach to morpho-syntactic parsing of agglutinative and other (highly) inflectional languages. In: *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*. pp. 261–268. ACL '99, Association for Computational Linguistics, College Park, Maryland (1999)
9. Redd, T.K., Read-Brown, S., Choi, D., Yackel, T.R., Tu, D.C., Chiang, M.F.: Electronic health record impact on productivity and efficiency in an academic pediatric ophthalmology practice. *Journal of AAPOS* 18(6), 584–589 (2014)
10. Sager, N., Lyman, M., Bucknall, C., Nhan, N., Tick, L.J.: Natural language processing and the representation of clinical data. *Journal of the American Medical Informatics Association* 1(2) (Mar/Apr 1994)
11. Siklósi, B., Novák, A.: Identifying and clustering relevant terms in clinical records using unsupervised methods. In: Besacier, L., Dediu, A.H., Martín-Vide, C. (eds.) *Statistical Language and Speech Processing*, pp. 233–243. *Lecture Notes in Computer Science*, Springer International Publishing (2014)
12. Siklósi, B., Novák, A.: A magyar beteg. X. Magyar Számítógépes Nyelvészeti Konferencia pp. 188–198 (2014)
13. Siklósi, B., Novák, A.: Restoring the intended structure of hungarian ophthalmology documents. In: *Proceedings of the 2015 Workshop on Biomedical Natural Language Processing*. Association for Computational Linguistics (2015)
14. Siklósi, B., Novák, A., Prószéky, G.: Context-aware correction of spelling errors in Hungarian medical documents. *Computer Speech and Language in press*(0), – (2014), <http://www.sciencedirect.com/science/article/pii/S0885230814000795>

15. Siklósi, B., Novák, A., Prószéky, G.: Resolving abbreviations in clinical texts without pre-existing structured resources. In: Fourth Workshop on Building and Evaluating Resources for Health and Biomedical Text Processing. ELRA (2014)
16. Siklósi, B.: Clustering relevant terms and identifying types of statements in clinical records. In: Gelbukh, A. (ed.) Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science, vol. 9042, pp. 619–630. Springer International Publishing (2015)