

ИЗГРАЖДАНЕ НА НАЦИОНАЛЕН РЕГИСТЪР НА БОЛНИТЕ ОТ ЗАХАРЕН ДИАБЕТ

Д. Чаръкчиев, С. Захариева, Г. Ангелова, Св. Бойчева, Ж. Ангелов, П. Маринова, Г. Ненчовска, Л. Манева, А. Величков, Г. Петрова, К. Коприварова, И. Стоева, М. Боянов, Р. Савова, Р. Радев, Л. Стойкова-Чорбанова, Е. Ташева, П. Денчева, Е. Фотева, К. Славчева, Б. Стоянов, А. Стоев, С. Алексиева, Е. Котова, И. Ковачева, Т. Томов

BUILDING A BULGARIAN NATIONAL REGISTRY OF PATIENTS WITH DIABETES MELLITUS

D. Tcharaktchiev, S. Zacharieva, G. Angelova, S. Boytcheva, Z. Angelov, P. Marinova, G. Nentchovska, L. Maneva, A. Velitchkov, G. Petrova, K. Koprivarova, I. Stoeva, M. Boyanov, R. Savova, R. Radev, L. Stoykova-Tchorbanova, E. Tasheva, P. Dentcheva, E. Foteva, K. Slavtcheva, B. Stoyanov, A. Stoev, S. Alexieva, E. Kotova, I. Kovatcheva, T. Tomov

Резюме. Статията представя изграждането на национален регистър на болните от захарен диабет с помощта на информационни технологии за обработка на естествен език и големи масиви от клинични данни.

Summary. The article presents the building of national registry of patients with diabetes mellitus using the technologies for Natural Language Processing and Big Data analysis.

Key words: Diabetes registry, Natural Language Processing (NLP), Big Data

Въведение

Захарният диабет се превръща в световна пандемия. На 21 декември 2006 г. Общото събрание на ООН прие резолюция, която признава, че светът е заплашен от епидемично разпространение на захарния диабет. За първи път едно неинфекциозно заболяване се определя като световна заплаха. Самостоятелно или в съчетание с други рискови фактори (в рамките на т.нар. метаболитен синдром), по своята същност и последици, захарният диабет представлява сърдечно-съдова болест с огромна човешка и икономическа цена. Данни получени от българско мултицентрово транзверзално проучване, включващо 12 български региона и общо 2032 лица, съобразено с методологията на Международната диабетна федерация за оценка разпространението на диабета сред възрастни, оценява честотата на захарния диабет у нас на 9.6% (диагностициран – 7.1% и недиагностициран – 2.5%).

Информацията за заболяването е разпокъсана, не се структурира и индексира правилно по определени критерии и признаци, липсват достатъчно обхватни и достъпни бази данни, които да обслужват различни аспекти на лечебно-диагностичния процес и рехабилитацията на диабетиците. Не са налице възможности за извършване на задълбочени анализи и провеждането на адекватна здравна политика основана на доказателствата. В тази обстановка, със заповед на министъра на здравеопазването през м. октомври 2013 г. е определена работна група с ръководител проф. д-р Сабина Захариева, дмн, Национален консултант по ендокринология и болести на обмяната. Групата изготвя предложение за вида на данните, които са необходими за изграждане на регистър на болните от диабет в Република България, както и за организацията и реда за попълване на регистъра и необходимите нормативни промени за реализацията му. През месеците ноември и декември 2015 г. екип от специалисти по медицинска информатика и информационни технологии (Д. Чаръкчиев, Г. Ангелова, Св. Бойчева, Ж. Ангелов, П. Маринова, Л. Манева, А. Величков) създава и прилага система за псевдоанонимизация на клинични записи, разработва и прилага методология и програмно осигуряване за извличане и анализ на факти от медицински текст. Реализиран е прототип на системата на регистъра. От 2015 г. Министерството на здравеопазването възлага с договор на Университетската специализирана бол-

ница за активно лечение по ендокринология „Акад. Иван Пенчев” ЕАД (УСБАЛЕ) да поддържа медицински регистри, осъществявайки дейности по системно събиране, съхранение, анализ, интерпретация и публикуване на данни за лицата с диабет и предразположение към диабет и да поддържа база данни на болните от диабет.

Цел и задачи

Основната цел, която следва да се изпълни при изграждането на регистъра, е подобряване на общественото здраве и качеството на здравното обслужване на пациентите със захарен диабет.

Специфичните цели включват събиране, обобщаване и предоставяне данни на регионално и национално ниво за заболяемостта и болестността от захарен диабет в Р България; подпомагане на МЗ и НЗОК при адекватното планиране и осигуряване на лечение и медицински грижи на пациентите със захарен диабет. Регистърът следва да дава възможност за контрол на състоянието на конкретния пациент; да предоставя данни за извършване на оценка на ефективността на здравните грижи, както и за взимане на обосновани решения базирани на доказателства.

Задачите, които се решават включват определяне на обхвата и структурата на данните, които са необходими за изграждане на регистъра, както и организацията, периодичността и начина за тяхното събиране. Важна задача е изграждането на релевантна информационна система, даваща възможност за системно събиране, съхранение, анализ и интерпретация на данните. При изграждането на системата на регистъра следва да се използват само анонимни данни, като в НЗОК се прилага методика за псевдоанонимизация на данните от отчетите на общопрактикуващите лекари и специалисти.

Материал и методи

Анализирани са данните от всички амбулаторни посещения при общопрактикуващи лекари и специалисти за 2012, 2013 и 2014 година. При всеки преглед и консултация, независимо от техния повод, лекарите попълват амбулаторен лист. Амбулаторните листове се изготвят на хартиен носител, а данните се съхраняват от информационните системи на личните лекари и специалисти, съгласно изискванията (Бл. МЗ-НЗОК № 1). При ежемесечния отчет на дейността

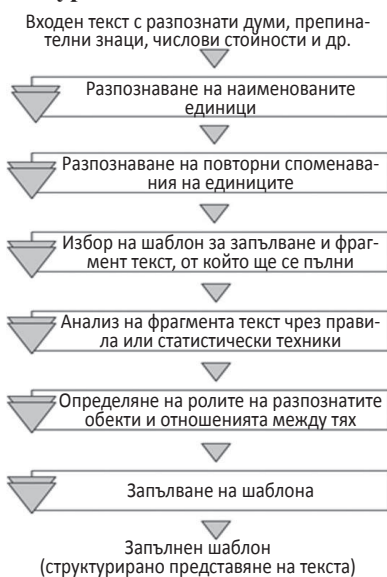
им амбулаторните листове се предават на НЗОК като файлове в XML формат. Файловете съдържат кодирани данни, касаещи идентификацията на пациента и лекаря, медико-административна информация за вида на прегледа, извършените дейности, издадените направления и документи и поставените диагнози. Данните от анамнезата, обективното състояние, изследванията и терапията се внасят като свободен текст в съответните полета. Някои системи генерират и записват в полето за описание на терапията предписаните частично или напълно реимбурсирани медикаменти, включително и кодовете им по номенклатура определена от НЗОК.

Анализиран са общо 112 412 712 посещения на пациенти при общопрактикуващи лекари и специалисти през периода от 01.01.2012 г. до 31.12.2014 г. Броят на пациентите, извършили посещения при лични лекари и специалисти по години е както следва: 2012 г. – 509 858 пациенти; 2013 г. – 5 166 202 пациенти; 2014 г. – 5 190 419 пациенти.

УСБАЛЕ има дългогодишен опит при създаването и работата със значителни масиви информация за диабетно болните, проследяването и контрола на тяхното състояние. Още в началото на 90-те години на миналия век са определени основните параметри на данните необходими за успешния контрол и определяне на подходящите лечебно-диагностични мероприятия при лицата със захарен диабет. Публикуван е клиничен минимум данни за контрол на тяхното състояние, който включва давност на диабета, тип, лечение, усложнения и придружаващи заболявания, антропометрични показатели (ръст, тегло, БМИ, обиколка на талията), витални показатели (артериално налягане), лабораторни показатели (кръвна захар на гладно и постпрандиално, хемоглобин А1С, Креатинин, Холестерол, HDL-холестерол, триглицериди и др), рискови фактори, обучение на пациента, болничен престой, вкл. и временна нетрудоспособност, бременности и раждания, качество на живот и др.

При анализа на електронната медицинска документация се установи, че над 90% от необходимата ни информацията се запазва във формата на неструктуриран текст. Това налага приложение за методи за обработка и анализ на естествен език (Natural Language Processing – NLP). Методите за анализ на български медицински текст са разработвани в продължение на над 8 години в рамките на национални и европейски проекти (ЕВТИМА, PSIP и др.) и са публикувани в български и чужди издания. Схема на класическа обработка на текст е показана на **Фигура 1**.

Фигура 1.



За идентификация, анализ и представяне на зависимости в големи масиви от данни е използвана системата *BITool* на фирма АДИСС Лаб ООД.

Резултати и обсъждане

Създадена е програмна система за съхраняване на данните, като информацията от всички визити на даден пациент е групирана в структуриран полицев запис (**Фигура 2**).

За болните, които не се прегледани от лекари, работещи с НЗОК е осигурена възможност за регистрация при тяхно съгласие от лекуващия ги лекар, която да се извършва с помощта на WEB базиран интерфейс и да включва ограничен брой показатели.

При организацията на събирането на данните за регистъра на болните със захарен диабет се спазват следните основни принципи:

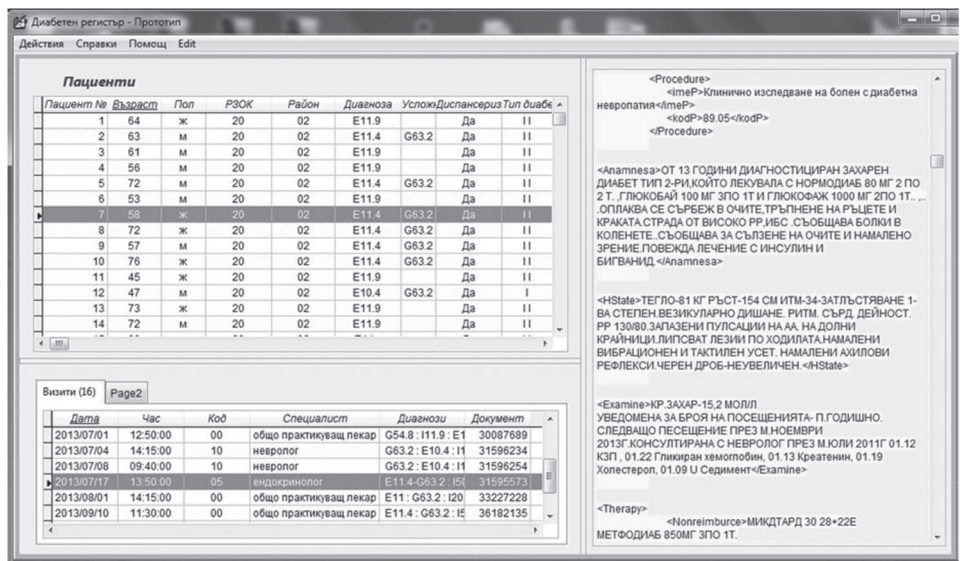
- ▶ не се променя организацията на работа на лекарите – данните се регистрират в стандартните информационни системи работещи в кабинетите на ОПЛ, специалистите и лечебните заведения и се предават по електронен път;
- ▶ използват се въведените от НЗОК XML формати за обмен на данните при спазване на всички законови изисквания за защита;
- ▶ практически не се отделя допълнително време и средства за допълнителна регистрация на данни и не се нарушава диагностично-лечебния процес;
- ▶ осигурява се еднократно въвеждане, многократно и многоаспектно използване на информацията;
- ▶ има възможност за саморегистрация и регистрация от лекари не работещи с НЗОК с помощта на WEB базиран интерфейс.

Следващите няколко примера илюстрират възможностите на системата и въведената организация на събиране на данните.

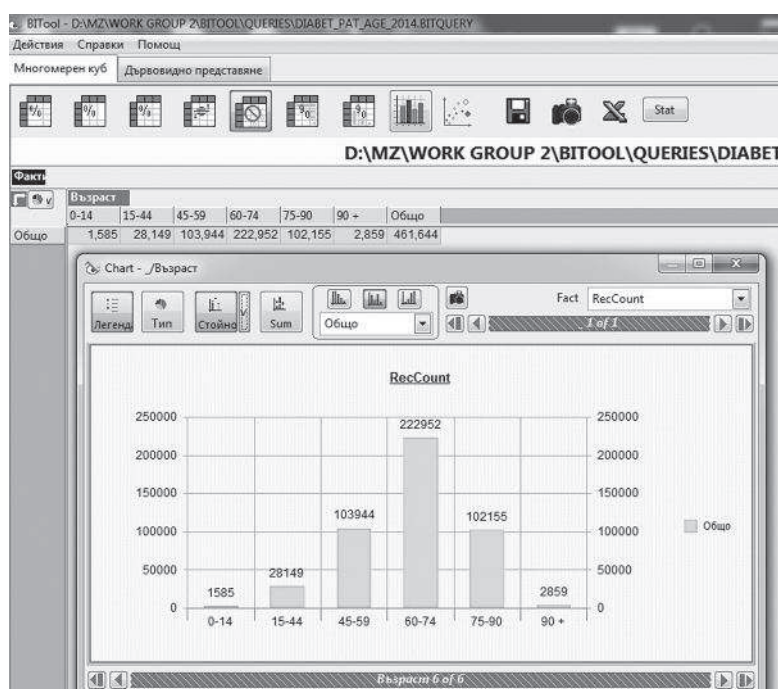
Общият брой на установените болни осигурени български граждани със захарен диабет по години е както следва: 2012 г. – 431 197; 2013 г. – 446 881 и 2014 г. – 461 644. (**Фигура 3**)

По данни на НЗОК броят на пациентите регистрирани през 2013 г. с рецептурни книжки със заверени диагнози захарен диабет и неговите усложнения е 384857. С помощта на високоэффективните технологии за извличане и анализ на факти от медицински текст, броят на установените болни от захарен диабет е с 96 983 болни повече от регистрираните в базите данни на НЗОК болни.

Фигура 2. Интерфейс на програмната система със структуриран здравен запис



Фигура 3. Разпределение на болните по възраст (2014 г.)



Установено е неравномерна разпределение на болните от захарен диабет по области – през 2014 г. процентът на диабетичите от обръналите се за преглед осигурени лица в Монтана е бил 10.63%, в София-град 8.95% в Благоевградска област 6.92%, в Кърджали 6.87%.

Важно е да се анализира компенсацията на диабетно болните, като функция от тяхната диспансеризация, с оглед да се оцени ефективността на комплекса от диагностично-лечебни мероприятия, осъществявани при диспансерното наблюдение. Анализът включва 359 682 болни, наблюдавани през тригодишния период 2012-2014 г., за които има необходимите данни от измерванията на кръвната захар и гликирания хемоглобин. Критерий за разделяне на болните на компенсирани и декомпенсирани е нивото на гликирания хемоглобин съответно под 7% и над 7%. При липса на данни за гликирания хемоглобин в текстовия блок за описание на изследванията е анализирано нивото на кръвната захар на гладно и два часа след нахранване. Критерии за компенсация – кръвна захар на гладно – под 7.8 ммол/л и кръвна захар след нахранване – под 10 ммол/л. Тъй като периодът на оценка на компенсацията е продължителен (тригодишен), се отчита времето, през което болният е бил компенсирани, като процент от целия период на наблюдение. Има се в предвид също, че гликираният хемоглобин дава информация за компенсацията в период на до 3 месеца преди извършване на изследването. Пациентът се счита за компенсирани и за последващия период, до следващото измерване. Можем да разделим болните на 2 групи – пациенти, които през по-голямата част от периода са били компенсирани (над 50% от тригодишния период) и които са били декомпенсирани през над 50% от периода.

Процентът на компенсирани болни със захарен диабет при диспансеризираните лица е 23.48% а при недиспансеризираните лица е по-висок и достига 34.66%.

Резултатът дава основание да се обсъди ефективността на провежданите диагностично-лечебни дейности в рамките

Фигура 4. Успеваемост при извличане на информация от български клинични текстове

	Лекарства	Лабораторни тестове
Брой извадени стойности от 6200 епикризи	160 892	114 441
Точност	97.28%	98.20%
Покриваемост	99.59%	99.99%
Средно F	98.42%	99.04%

на диспансеризацията и изчислен за всеки общопрактикуващ лекар поотделно може да се използва в комплексната оценка на дейността на лекаря.

За оценка на успеваемостта на приложената методика и софтуерни модули за обработка на текст с цел извличане на информация (IE) се прилагат следните мерки: *Precision* (точност – % коректно извлечени единици спрямо всички извлечени от тестовия корпус); *Recall* (покриваемост – % коректно извлечени единици спрямо всички налични в тестовия корпус); $F = 2 * Precision * Recall / (Precision + Recall)$. Например *F* за английски новинарски текстове достига над 95% при разпознаване на наименовани единици и ~80% при разпознаване ролите на обектите.

Нашата успеваемост при извличането на информация от клинични текстове (Фигура 4) непрекъснато се подобрява и надхвърля получените резултати от опитни кодировници.

Заклучение

Извършените анализи на големи масиви от данни (Big Data) включващи 112 412 712 посещения на пациенти при общопрактикуващи лекари и специалисти от доболничната помощ, недвусмислено демонстрират възможностите на съвременните технологии за прецизно извличане на информация от медицински текстове и предоставяне на необходимите данни за изграждането и поддържането на национален регистър на болните от захарен диабет.

Благодарности

Авторите благодарят на МЗ, НЗОК, УСБАЛЕ, МУ – София и АДИСС Лаб ООД за оказаната подкрепа. Изследванията, представени в тази статия, са частично подпомогнати от проекта AComIn “Advanced Computing for Innovation” FP7 grant 316087 Capacity Programme in 2012 – 2016.

Библиография

1. Борисова А-М., А. Шинков, Й. Влахов и сътр. Разпространение на захарния диабет и преддиабет в България днес. Ендокринология 4, 2012.
2. http://www.medicalnet-bg.org/usbale_old/doc/diabet.pdf
3. Чаръкчиев Д., Св. Бойчева, Г. Ангелова, И. Николова. Електронни здравни записи и автоматизирана обработка на медицински текст, Шумен, Инкома, 2012, 61 с.
4. Tcharaktchiev D., G. Angelova, S. Boytcheva, Z. Angelov, and S. Zacharieva. Completion of structured patient descriptions by semantic mining. In: Koutkias V., J. Niès, S. Jensen, N. Maglaveras, and R. Beuscart (Eds.), Patient safety informatics - adverse drug events ..., IOS Press, Studies in Health Technology and Informatics series, v. 166, 2011, 260 -269.
5. Boytcheva S., D. Tcharaktchiev, G. Angelova. Contextualization in automatic extraction of drugs from Hospital Patient Records. In A. Moen et al. (Eds) User centred networked health care, IOS Press, Studies in Health Technology and Informatics series, v. 169, 2011, 527-531.
6. PSIP Project Patient Safety through Intelligent Procedures in Medication, FP7-ICT-2008, № 216130. Deliverable 2.
7. Иванов Л., Ганова-Йоловска М, Константинов Б. Качеството на кодиране и достоверност на медицинската информация за разпределение на финансовите ресурси ..., Соц. Мед, 1999; 4, 32-34.

Автор

Димитър Чаръкчиев, Медицински университет – София, dimitardt@gmail.com