

STRUKTURÁLNÍ KVALITA DAT V E-HEALTH

Marek Mateják

V e-Health se často mluví o velkém potenciálu nestrukturovaných dat často v podobě volných textů lékařských zpráv a anamnéz. Z principu je tento potenciál však vždy menší než u zpracování těch samých dat ve strukturované podobě. Důvodem je, že strukturovaná data mají lépe definovaný odborný význam daný jejich strukturou. Podobně mají stejný význam pro autora i nestrukturovaná data, protože jim podvědomě svou vlastní strukturu přiřazuje. Bohužel výklad dané struktury a ve vzácném případě dokonce i vět nebo i termínů se může u různých odborníků lišit, pokud není standardizován. Proto se naskytuje otázka, jak zlepšit strukturální kvalitu těchto dat v e-Health. Ideálně již při jejich vzniku.

Úvod

Z pohledu počítačového zpracování dat je dnes nejlepší možnou elektronickou reprezentací dat databáze s vhodně zvoleným normalizovaným datovým modelem. Dle teorie, normalizace datových modelů nesmí při jejich definování záležet na konkrétní implementaci (tj. zda databáze bude SQL, objektová, NoSQL, NewSQL nebo jiná). Naopak, datový model by měl zahrnovat význam a kontext všech popisovaných dat [1]. Ideální cesta je, pokud je daný datový model již standardizován a dobře definován. Bohužel trend je takový, že se spíše definuje a standardizuje rozhraní na výměnu zdravotních a zdravotnických dat [2]. V mnoha případech jsou však tyto standardy navrženy už částečně v normalizovaném tvaru. To umožňuje převzetí jejich atributů i vztahů do datamodelu téměř beze změny [3].

V praxi jsou však zdravotničtí pracovníci zvyklí psát informace do jednoho textu, který většinou nazvou anamnézou nebo propouštěcí zprávou. Takto nestrukturovaný text je stále ještě možné alespoň částečně strukturovat, ale i když zapojíme umělou inteligenci a hodně sofistikované dobývání znalostí, tak se jen horko těžko dostáváme do stavu, který by zaručil data zadané zdravotnickým pracovníkem již v strukturované podobě.

Otázkou však je, je-li lepší učit počítače číst a poslouchat data nebo uživatele učit data zapisovat a diktovat. Možná nejlepší způsob je obě cesty zkombinovat a již v průběhu psaní/mluvení uživatele navigovat nebo ho upozorňovat na počítačem strukturovanou podobu jeho textu. K těmto nástrojům se hodí rozsáhlé slovníky termínů a jejich závislosti jako je SNOMED CT, LOINC, ICD a jiné. Bohužel ne všechno je lokalizované do češtiny, a proto je použití pro český jazyk značně omezeno.

Metody

Příkladem rozhraní na přenos zdravotních a zdravotnických dat je HL7 standard FHIR, kde disjunktí návrh jednotlivých entit a jejich atributů na objektové úrovni umožňuje strukturovaně uložit data. Je dobré zdůraznit, že definice standardu je implementovaná pro JSON, XML i Turtle formát dat. Snadno je ji dokonce možné implementovat i pro jiné formáty dat, či dokonce i databázová schémata. Pro přijetí FHIR jako datamodelu je nutné:

1. Určit interní typ identifikátorů jednotlivých záznamů, který bude stejný jako interní typ každé reference na záznam. Reference jsou totiž cizími klíči, které musí být v databázových systémech stejného typu a stejných hodnot jako klíče primární, na které ukazují.
2. Odstranit nejednoznačnosti zápisu dat. Například záznam pro pozorování (Observation) umožňuje částečně pozorování sloučit do jednoho celku pomocí referencí Observation.hasMember nebo Observation.compo-

nent. Vzhledem k tomu, že existují případy, kdy není možné jednoznačně určit, kterým z těchto způsobů data shlukovat, tak je vhodné v datamodelu komponenty reprezentovat pomocí referencí. Což ulehčí vyhledávání i zpracování těchto dat. V případě požadavku je možné vybrané shluky zpětně mapovat na komponenty.

3. Sjednotit a případně i mapovat hlavičkové údaje každého záznamu. Jako jsou identifikátory pacienta, autora, organizace autora, práva, souhlasy, čas vytvoření záznamu a čas poslední modifikace záznamu. Pod mapováním se myslí vypuštění daného atributu z FHIR, který lze odvodit z hlavičky záznamu. Datamodel tak neukládá tu samou informaci na více místech a namísto toho bezprostředně po načtení záznamu z databáze tuto informaci odvodí.

Data uložená pod takto zvoleným datamodelem s klíčovými hodnotami ve formě kódů z mezinárodních číselníků je možné zpracovávat velmi efektivně. Pomocí indexů a různých efektivních algoritmů v databázových prostředích jsou dotazy na data zpracovány řádově v desetinách sekundy i při stovkách milionů záznamů a několika terabajtových databázích. Ať už se data ukládají centrálně nebo distribuovaně, jejich zpracování a přístupnost v reálním čase jsou technologicky možné prakticky bez jakéhokoliv čekání, pokud mají vhodnou strukturu. Najít něco v datech, které vhodnou strukturu nemají, je však obvykle zdoluhavý problém. Ano, je možné indexovat i taková data a pomocí speciálních algoritmů v datech hledat pomocí regulárních výrazů. Výsledek je však nejistý a je třeba ho manuálně projít. Proto je otázkou, jestli lze tento postup aplikovat i pro jiné zpracování, než je vyhledávání.

Výsledky

V některých případech je možné na základě hodnot odhadovat atributy a typy FHIR záznamů i číselníkových kódů. Proto je možné interaktivně již při zadávání předvyplnit dané FHIR záznamy jenom na základě rozpoznání termínů z volného textu nebo řeči. Tyto záznamy však uživatel musí mít možnost upravit a schválit. Ve výsledku to uživatele donutí při dalším zadáváním používat strukturu bližší a bližší FHIR i termínům číselníků tak, aby následných úprav bylo co nejméně. Zároveň se může při zadávání učit i rozpoznávací algoritmus, protože má zpětnou vazbu.

V dnešní době je dobré si uvědomit, že zadávání může být hlasové a i zpětné prezentování dat může být snadno syntetizováno do hlasu. Což uvolní zdravotnickým pracovníkům ruce i oči na čas, který mohou investovat na kvalitnější kontakt s pacientem. Dokonce to celé může v extrémním případě běžet jenom jako aplikace na mobilním telefonu, která tyto data posílá i načte do/z nemocničního informačního systému.

Diskuze

Informační technologie jsou dobrý sluha, ale zlý pán. A proto je nutné často zvážit, zda zavádění daného produktu je skutečně win-win řešením pro všechny zúčastněné strany. Nikdo nechce trávit hodiny zbytečným přepisováním něčeho, co se stejně možná ani dál k ničemu nepoužije. Zároveň se těžko opouští zaběhnuté koleje každodenního života, kde se rituály práce s kartotékou záznamů již zdají být nenahraditelné. Proto je nutné produkty dotáhnout do formy a stavu, kdy budou nejenom ulehčovat práci zdravotnickým pracovníkům, ale zároveň budou bezpečným řešením pro pacienty [4], pro různé typy výměny zdravotnické dokumentace, pro záchranné služby, pro komunikaci s odbornými vyšetřeními, s hospitalizacemi, s pojišťovkami, s e-Receptem, s e-Neschopenkou, atd. [5]

Literatura

- [1.] M. Mateják, J. Potůček, a J. Kofránek, „NOVÁ GENERÁCIA ELEKTRONICKÝCH ZDRAVOTNÝCH ZÁZNAMOV“, *Sborník Příspěvků MEDSOFT 2016*, roč. 2016.
- [2.] Marek Mateják, Libor Seidl, Karel Zvára, „IHE profily“, *Sborník Příspěvků MEDSOFT 2019*, roč. 2019.
- [3.] Libor Seidl a Michal Potůček, „UŽITÍ REFERENČNÍCH MODELŮ VE ZDRAVOTNICKÉ INFORMATICE“, *Sborník Příspěvků MEDSOFT 2018*, roč. 2018.
- [4.] Marek Mateják, Libor Seidl, a Michal Potůček, „DEFINICE SOUKROMÍ PACIENTA V ELEKTRONICKÝCH ZDRAVOTNÍCH ZÁZNAMECH“, *Sborník Příspěvků MEDSOFT 2018*, roč. 2018.
- [5.] O. Heinze, M. Birkle, L. Köster, a B. Bergh, „Architecture of a consent management suite and integration into IHE-based regional health information networks“, *BMC Med. Inform. Decis. Mak.*, roč. 11, č. 1, s. 58, 2011.

Kontakt:

Mgr. Marek Mateják, Ph.D.

e-mail: marek@matfyz.cz