Univerzita Karlova v Praze
1. lékařská fakulta

Autoreferát disertační práce

Jazyk lékařských zpráv a jeho informačně lexikální analýza

Mgr. Petra Přečková

2011

Doktorské studijní programy v biomedicíně

Univerzita Karlova v Praze a Akademie věd České republiky

Obor: Biomedicínská informatika

Předseda oborové rady: Prof. RNDr. Jana Zvárová, DrSc.

Školicí pracoviště: Ústav informatiky AV ČR, v.v.i., Oddělení medicínské informatiky

Školitel: Prof. RNDr. Jana Zvárová, DrSc.

Disertační práce bude nejméně pět pracovních dnů před konáním obhajoby zveřejněna k nahlížení veřejnosti v tištěné podobě na Oddělení pro vědeckou činnost a zahraniční styky Děkanátu 1. lékařské fakulty.

Poděkování

Tato práce byla podpořena následujícími projekty: LN00B107 Ministerstva školství, mládeže a tělovýchovy ČR; 1ET200300413 Akademie věd ČR; AV0Z10300504 Ústavu informatiky AV ČR, v.v.i. a 1M06014 Ministerstva školství, mládeže a tělovýchovy ČR.

Velké poděkování patří mojí školitelce, Prof. RNDr. Janě Zvárové, DrSc., za její neocenitelné rady a pomoc během celého mého doktorského studia, ale i všem kolegům z oddělení medicínské informatiky Ústavu informatiky AV ČR, v.v.i. a v neposlední řadě mojí rodině za její vytrvalou velkou podporu.

Obsah

Abstrakt

Cílem disertační práce byla informačně lexikální analýza českých lékařských zpráv a využitelnost mezinárodních klasifikačních systému v českém zdravotnickém prostředí. Analýza lékařských zpráv byla založena na atributech Minimálního datového modelu pro kardiologii (MDMK). Byly použity lékařské zprávy psané volným textem a strukturované lékařské zprávy uložené v softwarové aplikaci ADAMEK. Pro práci byly využity zejména klasifikační systémy SNOMED CT a MKN-10. Bylo porovnáno, jak dobře jsou atributy MDMK zaznamenány v textových lékařských zprávách a v lékařských zprávách zaznamenávaných strukturovaně pomocí softwarové aplikace ADAMEK. Byla provedena jazyková analýza českých textových lékařských zpráv. Byla navržena nová aplikace pro měření diverzity lékařských zpráv psaných v jakémkoli jazyce. Tato nová aplikace je založena na obecných konceptech diverzit a byla odvozena z f-diverzity, relativní f-diverzity, vlastní f-diverzity a marginální f-diverzity. Závěrem práce je zjištění, že zapisování lékařských zpráv formou volného textu je velice nesourodé a není standardizováno. Použití standardizované terminologie by přineslo výhody lékařům, pacientům, administrátorům, softwarovým vývojářů a plátcům a pomohlo by poskytovatelům zdravotnické péče tím, že by poskytovalo kompletní a snadno dostupné informace, které náleží k procesu péče o zdraví a to by vedlo k lepší péči o zdraví. Použití mezinárodních klasifikačních systémů je nezbytným prvním krokem, který umožní sémantickou interoperabilitu heterogenních elektronických zdravotních záznamů.

Klíčová slova: *terminologie, synonyma, klasifikační systémy, tezaurus, nomenklatura, elektronický zdravotní záznam, sémantická interoperabilita, kardiologie, diverzita*

Abstract

The objective of the dissertation thesis has been the information-lexical analysis of Czech medical reports and the usability of international classification systems in the Czech healthcare environment. The analysis of medical reports has been based on the attributes of the Minimal Data Model for Cardiology (MDMC). Narrative medical reports and structured medical reports from the ADAMEK software application have been used. For the thesis SNOMED CT and ICD-10 classification systems have been used. There has been compared how well attributes of MDMC are recorded in narrative and structured medical reports. The language analysis of the Czech narrative medical reports has been made. A new application for measuring diversity in medical reports written in any language is proposed. The application is based on the general concepts of diversities derived form f-diversity, relative f-diversity, self f-diversity and marginal f-diversity. The thesis has come to the conclusion that using a free text in medical reports is not consistent and not standardized. The standardized terminology would bring benefits to physicians, patients, administrators, software developers and payers and it would help healthcare providers as it could provide complete and easily accessible information that belongs to the process of health care and it would result in better health care. The use of international classification systems is a necessary first step to enable semantic interoperability of heterogeneous electronic health records.

keywords: *terminology, synonyms, classification systems, thesaurus, nomenclature, electronic health record, semantic interoperability, cardiology, diversity*

1. ÚVOD

Styl zapisování textových lékařských zpráv není v České republice nijak standardizován. Stejně tak i vymezení, pojmenování a třídění lékařských pojmů není optimální. Dokladem je skutečnost, že pro jeden pojem existuje často více než deset synonym. Praktickým negativním důsledkem v lékařství je situace, kdy je například efekt nového léku nebo hodnot nové vyšetřovací metody u dané diagnózy popisován ve dvou publikacích. Pokud je chápání této diagnózy v každé z uvedených publikací poněkud posunuto a jedná se tedy o rozdílné množiny pacientů, můžeme se často setkat i s kontroverzními výsledky, což hodnotu výsledné informace samozřejmě snižuje.

Se zaváděním výpočetní techniky v lékařství se tento problém prohloubil, neboť její využívání předpokládá větší jednoznačnost zadávání dat, vymezení pojmů a jejich přesné pojmenování.

Obecně je velmi výhodné využívat v odborné terminologii pro jeden pojem vždy pouze jediný výraz. Synonyma lze sice počítač naučit, zvětšují však rozsah slovníku databáze i počet nezbytných operací, což prodlužuje komunikaci. Synonymie v odborné terminologii vede při sdělování informací navíc k nepřesnostem a nedorozumění. V současné lékařské terminologii se lze setkat s řadou synonym pro jediné onemocnění. Z tohoto důvodu začaly vznikat kódovací systémy, které rychle poskytnou kód pro libovolný biomedicínský poznatek.

V současné době dochází k velkému rozvoji elektronických zdravotních záznamů. Existuje obecná shoda, že elektronické lékařské zprávy mají potenciál zlepšit kvalitu lékařské péče [1]. V konceptu elektronického zdravotního záznamu je pacient chápán jako aktivní partner, který má přístup ke svým zdravotním datům[2].

2. CÍLE PRÁCE

Jelikož v současné době existuje ve zdravotnictví více než 100 různých klasifikačních systémů, jedním z hlavních cílů práce bylo zmapování těchto klasifikací a vybrání těch nejvhodnější pro potřeby českého zdravotnictví.

Bezpečná a vhodná výměna klinických informací mezi různými elektronickými zdravotními záznamy je nezbytná k zajištění kontinuity péče o pacienty a to v různých časech, na různých místech a u různých poskytovatelů zdravotní péče. Mapování atributů elektronických zdravotních záznamů na mezinárodní klasifikační systémy je tedy důležitým krokem pro sémantickou interoperabilitu mezi těmito různými systémy a bylo součástí této vědecké práce. Elektronické zdravotní záznamy a sémantická interoperabilita jsou velice aktuálními problémy a jsou diskutovány v mnoha článcích [3], [4], [5], [6]. Sémantická interoperabilita založená na českém jazyce byla zkoumána v [7], [8], [9], [10], [11].

Klinické údaje z elektronických zdravotních záznamů tradičně obsahovaly malé množství strukturovaných dat (často získané ze seznamu možných odpovědí) a větší množství volného textu [12]. Ve své práci jsem se zaměřila na oba dva druhy těchto klinických údajů.

Velká část práce je věnována Minimálnímu datovému modelu pro kardiologii vytvořeného v rámci výzkumného centra EuroMISE – Kardio, jeho rozboru a mapování jeho atributů na mezinárodní klasifikační systémy.

Cílem práce byla také lexikální analýza lékařských zpráv psaných volným textem, jejich jazykový rozbor a analýza rozdílnosti zapisování mezi různými lékaři.

Tato disertační práce také ukazuje novou aplikaci pro měření diverzity lékařských zpráv psaných v jakémkoli jazyce. Tato nová aplikace je založena na obecných konceptech diverzit a byla odvozena z f-diverzity, relativní f-diverzity, vlastní f-diverzity a marginální f-diverzity. Byly porovnány diverzity vybraných atributů v textových lékařských zprávách a

porovnány výsledky s diverzitami stejných atributů ve strukturovaných lékařských zprávách a získány vědecky zdůvodněné závěry pro posouzení informace obsažené v těchto dvou typech různých zpráv.

3. MATERIÁL A METODIKA

3. 1. Kódovací a klasifikační systémy

Kódovací systémy omezují variabilitu vyjadřování. Lze používat pouze schválené termíny a jejich spojení a to podle přesně stanovených pravidel. Obvykle jsou namísto schválených termínů používány formální kódy. V mnoha případech je užitečné, když kódovací systém rovněž ukazuje neschválené termíny, které jsou užívány jako synonyma pro schválené termíny. Kódovací systém, který je vybaven ještě takovou terminologickou informací, se nazývá thesaurus.

MKN – Mezinárodní klasifikace nemocí a přidružených zdravotních problémů

Mezinárodní klasifikace nemocí a přidružených zdravotních problémů (MKN) [13], [14], [15], [16] je českým překladem International Classification of Diseases and Related Health Problems (ICD). Jedná se o klasifikaci kódující lidská onemocnění, příčiny smrti, zdravotní problémy a další příznaky. MKN se používá k převodu diagnóz nemocí a jiných zdravotních problémů ze slovní podoby do alfanumerického kódu. Její základ byl položen již v roce 1893 [13] při klasifikaci příčin úmrtí s cílem umožnit mezinárodní porovnání. V roce 1948 převzala tuto klasifikaci Světová zdravotnická organizace WHO (World Health Organisation) a rozšířila ji o další diagnózy. Postupně tak začala vznikat všestranná pomůcka pro řízení zdravotnické politiky a pro výkaznictví ve vztahu ke zdravotnickým pojišťovnám a obdobným platebním systémům. Obsah MKN umožňuje systematické zaznamenávání, analýzu, výklad a porovnávání dat o úmrtnosti a nemocnosti, která jsou shromážděna v různých zemích nebo oblastech a v rozdílném čase.

SNOMED CT

SNOMED Clinical Terms [17], [18], [19], [20] vzniknul spojením dvou terminologií: SNOMED RT a Clinical Terms Version 3 (Read Codes CTV3). SNOMED RT znamená Systematized Nomenclature of Medicine Reference Terminology, kterou vytvořila College of American Pathologists. Slouží jako společná referenční terminologie pro shromažďování a získávání zdravotnických dat zaznamenaných organizacemi nebo jednotlivci. Clinical Terms Version 3 (Read Codes CTV3) vznikla v United Kingdom‘s National Health Service v roce 1980 jako mechanismus pro ukládání strukturovaných informací o primární péči ve Velké Británii.

V roce 1999 se tyto dvě terminologie spojily a vznikl tak SNOMED CT, což je vysoce komplexní terminologie skládající se z 19 hierarchií. Na jejím vytváření se podílí kolem 50 lékařů, sester, asistentů, lékárníků, informatiků a dalších zdravotnických odborníků. Byly vytvořeny speciální terminologické skupiny pro specifické terminologické oblasti jako je například ošetřovatelství nebo farmacie. SNOMED CT zahrnuje 311 000 aktivních konceptů, 794 000 anglických popisů a synonym a 920 000 sémantických vztahů. V roce 2007 přešla všechna práva SNOMED CT na International Health Terminology Standards Development Organisation (IHTSDO) sídlící v Dánsku.

V současné době existuje americká, britská, španělská a německá verze SNOMED CT.

MeSH

Medical **S**ubject **H**eadings [21], [22] je slovník kontrolovaný Národní lékařskou knihovnou (NLM) v USA. Tvoří ho skupina pojmů, které hierarchicky pojmenovávají klíčová

slova a tato hierarchie napomáhá při vyhledávání na různých úrovních specifičnosti. Klíčová slova jsou uspořádána jak abecedně tak hierarchicky. Na nejobecnější úrovni hierarchické struktury jsou široké pojmy jako např. „anatomie“ nebo „mentální onemocnění“. Hierarchie je 11stupňová. NLM využívá MeSH k indexování článků ze 4800 světových předních biomedicínských časopisů pro databázi MEDLINE/PubMED®. MeSH se využívá také pro databázi katalogizující knihy, dokumenty a audiovizuální materiály. Každý bibliografický odkaz je spojován se skupinou termínů v klasifikačním systému MeSH. Vyhledávací dotazy používají také slovní zásobu z MeSH, aby našly články na požadované téma. Specialisté, kteří MeSH slovník vytvářejí, ho průběžně aktualizují a kontrolují. Sbírají nové pojmy, které se začínají objevovat ve vědecké literatuře nebo ve vznikajících oblastech výzkumu, definují tyto pojmy v rámci obsahu existujícího slovníku a doporučují jejich přidání do slovníku MeSH.

LOINC

Klasifikační systém **L**ogical **O**bservations **I**dentifiers **N**ames and **C**odes (LOINC) [23], [24] je klinickou terminologií důležitou pro laboratorní testy a laboratorní výsledky. V roce 1999 byl LOINC přijat organizací HL7 jako preferované kódování pro názvy laboratorních testů a klinických pozorování. LOINC databáze obsahuje kategorie jako je chemické složení, hematologie, sérologie, mikrobiologie (včetně parazitologie a virologie) a toxikologie. Dále sem patří kategorie pro léky a krevní obraz. Klinická část obsahuje například kategorie jako záznamy pro životní funkce (puls, teplota, pravidelnost dýchání, krevní tlak), hemodynamiku, EKG, porodnický ultrazvuk, echo srdce, urologické snímání, gastroendoskopické procesy a další klinická pozorování.

ICD-O

Klasifikační systém ICD-O [25], [26] je rozšířením Mezinárodní klasifikace nemocí (ICD) pro kódování onkologie, která byla prvně publikována Světovou zdravotnickou organizací v roce 1976. Jedná se o čtyřdimenzionální systém, mezi jehož dimenze patří topografie, morfologie, průběh a diferenciace. Dimenze jsou určeny pro třídění morfologických typů nádorů. V současné době existuje její třetí verze.

TNM

TNM [27], [28] klasifikace je klinická klasifikace maligních nádorů, která se využívá pro účely srovnávání terapeutických studií. Vychází z poznatku, že pro prognózu onemocnění je zvláště důležitá lokalizace a šíření tumoru.

DSM

Mezi psychiatrické nomenklatury můžeme zařadit např. DSM (Diagnostic and Statistical Manual of Mental Disorder) [29], která obsahuje i definice jednotlivých pojmů. Jedná se o velice propracovanou nomenklaturu. Bohužel jde o uzavřený systém bez návaznosti na další obory lékařství.

3. 2. Konverzní nástroje

Rostoucí počet klasifikačních systémů a nomenklatur si vyžádal vytváření převodníků mezi hlavními systémy při přenosu informací mezi různými datovými bázemi. Nejrozsáhlejším projektem se stalo UMLS.

Cílem UMLS (Unified Medical Language System – Sjednocený systém medicínského jazyka) [30], [31], [32] vytvořeného v Národní lékařské knihovně (NLM) v USA, je usnadnění vývoje počítačových systémů, které se chovají jakoby „rozuměly“ významu

biomedicínského a zdravotnického jazyka. Za tímto účelem Národní lékařská knihovna vytváří a rozšiřuje UMLS Knowledge Sources (Znalostní zdroje UMLS) (databáze) a přidružené softwarové nástroje (programy), které mohou využívat systémový vývojáři při budování nebo zlepšování elektronických informačních systémů, které vytvářejí, zpracovávají, vyhledávají, integrují a/nebo shromažďují biomedicínská a zdravotnická data a informace. Tyto znalostní zdroje se dají využít také v informatickém výzkumu. Znalostní zdroje UMLS jsou záměrně víceúčelové. Nejsou optimalizované pro jednotlivé aplikace, ale mohou být aplikovány v systémech, které vykonávají několik funkcí zahrnujících jeden nebo více druhů informací, jako jsou např. záznamy o pacientech, vědecká literatura, doporučení a veřejná zdravotnická data. Přidružené softwarové nástroje UMLS pomáhají vývojářům při přizpůsobení a používání znalostních zdrojů UMLS pro konkrétní účely. Lexikální nástroje pracují efektivněji v kombinaci se znalostními zdroji UMLS, ale mohou být použity i samostatně.

Existují tři znalostní zdroje UMLS: Metathesaurus, Semantic Network a SPECIALIST lexicon. Jsou distribuovány s flexibilními lexikálními nástroji a instalačním programem MetamorphoSys, který umožňuje úpravy podle požadavků uživatele.

3. 3. Minimální datový model pro kardiologii

V rámci Centra biomedicínské informatiky navazujeme na výzkum z našich předchozích projektů. V letech 2000-2004 bylo jedním z cílů výzkumného centra EuroMISE – Kardio sestavení Minimálního datového modelu pro kardiologii (MDMK) [33], [34], [35].

Jelikož je kardiologie velice rozsáhlý obor, byl MDMK zaměřen pouze na aterosklerotická kardiovaskulární onemocnění. Cílem tohoto datového modelu bylo vytvoření minimálního souboru znaků, které je potřeba sledovat u pacientů z hlediska aterosklerotického kardiovaskulárního onemocnění, aby mohl být pacient následně zařazen mezi osoby nemocné či rizikové. MDMK se skládá z několika skupin znaků.

První část tvoří *administrativní údaje*, které jsou potřebné pro identifikaci pacienta. Další částí je *rodinná anamnéza*, zahrnující informace o matce, otci a libovolném počtu sourozenců. Dále následuje *sociální anamnéza a toxikománie*, která se zaměřuje na rodinný stav, fyzickou zátěž, psychickou zátěž, fyzické aktivity, míru kouření a míru požívání alkoholu. Část MDMK je věnována *alergiím* pacienta, zejména alergiím na léky. V části *osobní anamnézy* je zjišťována přítomnost diabetu mellitu, hypertenze, hyperlipoproteinémie, ischemické choroby srdeční a její konkrétní formy, je zjišťováno, zda pacient prodělal cévní mozkovou příhodu, zda se léčí s ischemickou chorobou periferních tepen, jsou zde atributy týkající se aneurysma aorty, ostatních relevantních chorob a u žen menopauzy. V části MDMK nazvané *Současné obtíže možného kardiálního původu* se lékaři zaměřují na dušnost, bolest na hrudi, palpitace, otoky, synkopu, kašel, hemoptýzu a klaudikaci. Další část MDMK zjišťuje, jakou *léčbu* pacient podstupuje, jaký má předepsaný druh diety a jaké užívá léky. V části *fyzikálních vyšetření* se zjišťuje pacientova hmotnost, výška, tělesná teplota, obvod boků, BMI, WHR, krevní tlak, tepová a dechová frekvence a patologické nálezy. *Laboratorní vyšetření* se zaměřují na glykémii, kyselinu močovou, celkový cholesterol, HDL-cholesterol, LDL-cholesterol a triaglyceroly. Poslední část MDMK tvoří *atributy vztahující se k EKG*, kde se zjišťuje rytmus, frekvence, průměrné intervaly PQ a QRS a je zde prostor pro celkový popis EKG.

Na základě MDMK byla vytvořena softwarová aplikace ADAMEK (**A**plikace **Da**tového **M**odelu **E**uroMISE centra – **K**ardio). Po jejím dokončení byl od března 2002 zahájen sběr dat v ambulanci preventivní kardiologie EuroMISE centra, která je spravována Městskou nemocnicí Čáslav. V současné době jsou v databázi ADAMEK zaznamenána anonymizovaná data o 1289 pacientech.

4. VÝSLEDKY

4. 1. Využití klasifikačních systémů pro sdílenou zdravotní péči

Mapování terminologie uváděné v aplikacích elektronického zdravotního záznamu na mezinárodně používané terminologické slovníky, tezaury, ontologie a klasifikace je základem pro interoperabilitu heterogenních systémů elektronického zdravotního záznamu. K zajištění interoperability však nestačí pouhé porozumění si na úrovni terminologických výrazů. Dalším předpokladem pro úspěšné sdílení dat mezi různými aplikacemi zdravotního záznamu je harmonizace klinického obsahu. Tato harmonizace nemusí být úplně stoprocentní, pak je ale možné sdílet pouze data, která jsou mezi aplikacemi společná. Interoperabilitu usnadní, pokud si odpovídají tzv. referenční informační modely jednotlivých aplikací zdravotních záznamů. Samozřejmě se nabízejí možnosti vzájemného mapování mezi těmito modely, což je však těžké vzhledem k odlišnému přístupu jednotlivých modelů.

Například HL7 RIM (Referenční informační model) [36] představuje model uzavřeného světa definovaného pomocí tříd, jejich atributů a vztahů mezi třídami. Pro další použití v konkrétní oblasti se od tohoto modelu odvozuje takzvaný D-MIM (Doménový informační model). Abychom se od takovéhoto modelu dostali ke zprávám nesoucím informace o zdravotním záznamu pacienta, použijeme tzv. R-MIM (*Refined Message Information Model*), který je podmnožinou D-MIM použitou pro vyjádření informačního obsahu jedné nebo více abstraktních struktur zpráv nazývaných též *Hierarchické popisy zpráv*.

Jiným příkladem je CEN TC 251, který definuje v evropském předběžném standardu ENV 13606 (Sdělování elektronických zdravotních záznamů, 4. část – zprávy pro výměnu informací) obsah elektronického zdravotního záznamu pomocí poměrně hrubého modelu specifikujícího 4 základní složky:

- *Folder* – popisující větší sekce záznamu daného subjektu,
- *Composition* – reprezentující jeden identifikovatelný příspěvek ke zdravotnímu záznamu daného subjektu,
- *Headed Section* – obsahující množiny údajů na jemnější úrovni než Composition,
- *Cluster* – identifikující skupiny údajů, které by měly zůstat seskupeny, hrozí-li ztráta kontextu.

Zcela jiný přístup používá asociace NEMA (National Electrical Manufacturers Association) při specifikaci DICOM SR (DICOM Structured Reporting), ve které dochází k rozšíření specifikace pro generování, prezentaci, výměnu a archivaci medicínských snímků DICOM na modelování celého zdravotního záznamu pacienta. Hlavní ideou zde je použít existující infrastrukturu DICOM pro výměnu strukturovaných zpráv, které představují hierarchický strom dokumentu s typovanými koncovými uzly. Sémantika jednotlivých uzlů je popsána kódovacími systémy jako např. ICD-10 či SNOMED.

Referenční model Synapses Object Model (SynOM) vytvořený v rámci projektu Synapses, resp. SynEx (Synergy on the Extranet) [37] je velmi podobný modelu definovanému v CEN ENV 13606. Jako typy sbíraných hodnot jsou zde využity tzv. archetypy – definice strukturovaně sbíraných údajů v určité doméně obsahující specifikovaná omezení zajišťující integritu celkového záznamu. Projekt dále pod záštitou neziskové openEHR Foundation pokračoval a definoval tzv. Good European Health Record (GEHR) [38]. V projektu odborníci specifikují požadavky elektronického zdravotního záznamu s hlavním cílem podpořit možnosti integrace a spolupráce heterogenních EHR aplikací. Za tímto účelem vznikl formální model specifikující GEHR architekturu (GEHR Object Model, GOM) a znalostní model specifikující klinickou strukturu záznamu pomocí archetypů.

Výstupy projektu openEHR lze v dnešní době považovat za významnou konkurenci standardům orientovaným na implementační aspekty EHR systémů.

Aby informatici mohli mapování na terminologie do budoucna využít, je vhodné na správnou terminologii myslet již od začátku, tj. jak při navrhování archetypů tak při tvorbě ostatních základních elementů v jiných typech modelů architektury zdravotních záznamů.

Jako příklad, jak správnou terminologii odkazovat již při vytváření archetypů, může fungovat editor od firmy *Ocean Informatics* [39]. Je možné přidat libovolný počet jazyků, ve kterých daný termín popíšeme. Zároveň je možné zvolit z dostupných terminologií ty, které použijeme k tomu, abychom definovali správný význam jednotlivých termínů. Na výběr máme více než 100 různých definovaných terminologií. Na dalších záložkách v tomto editoru definujeme jednotlivé termíny a na záložce Term bindings provedeme příslušné mapování našich termínů na termíny v terminologických slovnících.

Při analýze jsem zjistila, že přibližně 85 % atributů MDMK je obsaženo alespoň v nějakém klasifikačním systému. Většina z nich je obsažena v systému SNOMED CT.

K obdobnému závěru jsem dospěla při analýze možností standardizace atributů Datového standardu Ministerstva zdravotnictví České republiky (DASTA) [40]. Strukturované atributy v tomto standardu se však ve velké míře omezují na administrativní a laboratorní údaje. Při mapování administrativních údajů byly výsledky obdobné jako při mapování administrativních údajů v MDMK. Laboratorní údaje jsou v tomto standardu velmi podrobně specifikované pomocí tzv. Národního číselníku laboratorních položek [41].

V neposlední řadě jsem se zaměřila na mapování atributů vybraných klinických modulů komerčních nemocničních informačních systémů. Jako příklad uvedu výsledky mapování specializovaného EKG modulu v systému WinMedicalc. Vzhledem k velké specializovanosti tohoto modulu se podařilo namapovat přibližně 60 % atributů na různé klasifikační systémy. Převládající klasifikační problémy souvisí v tomto případě s příliš velkou granularitou atributů v tomto modelu (ejekční frakce 1, ejekční frakce 2, septum levé komory).

4. 2. Jazyk lékařských zpráv a využití mezinárodních klasifikačních systémů v minimálním datovém modelu pro kardiologii

Český jazyk patří k západní skupině slovanských jazyků. Mezi slovanskými jazyky patří do východní, nebo-li satémové, skupiny indoevropských jazyků. Čeština patří mezi jazyky s volným slovosledem [42].

Styl zapisování textových zpráv není nijak standardizován [43] a častou jsou psány formou volného textu [44]. Rozdíly najdeme nejenom ve zprávách od různých lékařů, ale i jednotliví lékaři často zapisují stejné koncepty v různých tvarech. Následující část je zaměřena na již zmíněné jazykové a lexikální, rozdíly v lékařských zprávách.

Diakritika: Někteří lékaři zapisují text bez použití diakritiky, např. „Bricho mekke nebolestive“. Většina z nich diakritická písmena ale používá.

Překlepy: Větším problémem jsou překlepy, které jsou velmi časté a text je potom dále velmi těžce použitelný pro počítačové zpracovávání.

Mezery: Podobnou záležitostí je i vynechávání mezer mezi slovy, kdy se ze dvou slov stává jedno slovo, jako například „pivopřestal“. Lékaři se různí v zapisování mezer před jednotkami. Můžeme se setkat jak s tvarem s mezerou, např. „2,5 mg“, tak i s tvarem bez mezery, např. „4mg“. Tak to je i s tvary, kde se používá lomítko. Někteří lékaři používají variantu bez mezer, např. „80/min“, jiní variantu s mezerami „70 / min“.

Číslice 0: Pro počítačové zpracování je také složité, když někteří lékaři používají místo číslice 0 velké písmeno O.

Zkratky: Jelikož lékaři mívají málo času na zapisování zpráv, dochází ke zkracování slov. Zkrácené tvary ale bývají různě dlouhé, například kyselina močová bývá zkracována jako kys. moč., kys. močová nebo KM. Může se stát, že v jedné a té samé zprávě je slovo zkráceno dvakrát a pokaždé jinak. Se zkrácenými tvary souvisí také to, že se setkáváme s vynecháním tečky za zkráceným slovem, např. „levostr kard insuf.“.

Zaokrouhlování: Další část, ve které můžeme nalézt mnoho rozdílů, souvisí s číselnými hodnotami. Zde se můžeme například setkat u stejného znaku u jednoho lékaře se zaokrouhlováním hodnot na celá čísla, u jiného lékaře s uváděním hodnot nezaokrouhlených, s přesností na jedno nebo dvě desetinná čísla. Někdy jsou číselné hodnoty znaku uváděné jako rozmezí, např. „70-80“. Častokrát bývá zadán pouze přibližný údaj, například „diastolický tlak kolem 70“. U některých znaků nejsou hodnoty vyjádřeny číslem, ale pouze slovně, např. „tlak je zcela v mezích normy“.

Římské a arabské číslice: Rozdíl je i v používání římských a arabských číslic. Například u zápisu o srdečních ozvách lze najít jak tvar „ozvy 2“, tak i „ozvy II“.

Synonyma: Český jazyk je velmi bohatý na synonyma a ta nacházíme i v lékařských zprávách. Jako příklad uveďme dolní končetiny versus nohy, hmotnost versus váha, iregulární versus nepravidelný, praktický lékař versus obvodní lékař versus prakt. lékař versus PL versus OL. Tepová frekvence bývá zapsána třemi různými způsoby: tep versus P versus fr. a mnoho dalších.

Pravopis: Někteří lékaři používají starší formy pravopisu, někteří novější, takže se můžeme setkat např. se znakem „cyanóza“, „cyanosa“, ale i „cyanoza“ nebo „hyperlipoproteinemie“, ale i „hyperlipoproteinémie“.

Časové údaje: Ani zaznamenávání časových údajů není sjednoceno. V lékařských zprávách se objevuje jak název měsíce, např. „únor 2006“, tak i pořadí měsíce, např. „2/2006“.

Podávání léků: Velmi odlišné je i zapisování rozpisu podávání léků. Stejná informace, kdy jedna tableta léku má být podávána ráno, bývá zapsána takto: 1 ráno, 1x ráno, 1-0-0, 1 tabl. ráno. Setkáváme se i s pouze slovním vyjádřením dávkování, jako například „jen zřídka“, „tabletu vezme až v poledne“, „denně“, „obd“, „příležitostně“, „při bolesti“, „dle hodnot QT“.

Hodnoty znaků: Často jsou stejné hodnoty znaku zapisovány řadou různých způsobů. Například:

- Hodnota znaku *diabetes mellitus* bývá zapsána jako: diabet, diabet., diabetes mellitus 2. typu, diabetička 2. typu na dietě, diabetes mellitus II. typu na dietě, DM 2.typu, DM 2. typu.
- *Dolní končetiny bez otoků* můžeme nalézt zapsané těmito způsoby: otoky DK nepozoruje, DK bez otoků, DK – bez otoků, DK neotékají, DK bez otoku, DK otoky 0. Přitom se jedná stále o tu samou informaci.
- Když hledáme v textových zprávách informace o *dušnosti*, najdeme tyto tvary: není dušná, není dušn, dušnost nepozoruje, dušnost neudává, bez dušnosti.
- Jak už jsme se dříve zmínili v souvislosti se synonymy, u *hmotnosti* bývají tyto informace: hmotnost 86 kg, V 86 kg, váha 86 kg, vaha 86 kg.
- Při studiu textových lékařských zpráv bylo nalezeno pět možností, jak bývá zapsáno, že je *srdeční akce pravidelná*: akce srd. prav., AS pravid., AS prav., akce pr. a cor- AS pravid.
- *Triacylglyceroly* bývají v textových lékařských zprávách zkracovány jako Tg, Tgl nebo TAG.

Nejedná se ale jenom o problém při zapisování lékařských zpráv, ale stejné chyby můžeme najít např. i na webových stránkách [45].

4. 3. Analýza znaků Minimálního datového modelu pro kardiologii v textových lékařských zprávách

V analýze 110 textových lékařských zpráv jsem vycházela ze znaků Minimálního datového modelu pro kardiologii. Lékařské zprávy byly anonymizované a z tohoto důvodu nebylo možné analyzovat administrativní data.

Podívejme se nyní, jak často byly zaznamenány vybrané znaky Minimálního datového modelu v lékařských zprávách psaných volným textem:

- *kouření* bylo zaznamenáno v 64,5 % textových zpráv (v 71 zprávách),
- *alergie* v 81, 8 % (v 90 zprávách),
- jestli pacient trpí *ischemickou chorobou srdeční* v 60,9 % (v 67 zprávách),
- přítomnost nebo nepřítomnost *dušnosti* byla zaznamenána v 71,8 % (v 79 zprávách),
- jestli pacienta trápí *bolest na hrudi* bylo zaznamenáno v 34,5 % (v 38 zprávách),
- otázky na *palpitaci* byly zaznamenány v 15,5 % (v 17 zprávách),
- odpovědi na atribut *otoky* byly nalezeny v 86,4 % (v 95 zprávách),
- *výška* byla zaznamenána v 85,5 % (v 94 zprávách)
- *diabetes mellitus* v 62,7 % zpráv (v 69 zprávách) a tak dále.

Nepříjemné u textových zpráv je fakt, že nevíme, zda vybraný znak, který nebyl ve zprávě zaznamenám, je z důvodu jeho nepřítomnosti nebo zda se na něj lékař nezeptal.

4. 4. Analýza znaků Minimálního datového modelu pro kardiologii v softwarové aplikaci ADAMEK

Jedním z cílů, na které se v poslední době v oblasti biomedicínské informatiky soustřeďuje stále větší úsilí, je vytvoření databázových systémů společně se softwarovými nástroji, které by mohly analyzovat získané údaje. A tak i po zformování Minimálního datového modelu pro kardiologii vyvstala přirozená potřeba sbírat data o pacientech v souladu s tímto modelem. Navíc, aby tato data byla dobře použitelná pro následné statistické či jiné zpracování a vyhodnocování, bylo žádoucí, aby tato data byla sbírána jednotným způsobem. Z tohoto důvodu byla vytvořena aplikace ADAMEK (Aplikace datového modelu EuroMISE – Kardio) [46], [47]. Byla sice snaha tvořit aplikaci ADAMEK jako systém, který by mohl sloužit pro vedení elektronické zdravotní dokumentace v kardiologických ambulantních zařízeních, ale není to její primární určení. Z tohoto důvodu v ní nejsou implementovány žádné funkce či nástroje výkaznictví pro zdravotní pojišťovny, statistické nástroje a řada dalších funkcí.

Celý záznam o pacientovi je rozdělen do části administrativa, rodinná anamnéza, sociální anamnéza, alergie, osobní anamnéza, obtíže, léčba, fyzikální vyšetření, laboratorní vyšetření a EKG.

Pro analýzu bylo využito 1119 lékařských zpráv z ambulance preventivní kardiologie EuroMISE centra.

Podívejme se nyní, jak často byly zaznamenány vybrané znaky Minimálního datového modelu ve strukturovaných lékařských zprávách uložených v softwarové aplikaci ADAMEK:

- *kouření* bylo zaznamenáno v 96,5 % (v 1080 zprávách),
- *alergie* v 95,2 % (v 1065 zprávách),
- jestli pacient trpí *ischemickou chorobou srdeční* v 93,4 % (v 1045 zprávách),
- přítomnost nebo nepřítomnost *dušnosti* byla zaznamenána v 93,6 % (v 1047 zprávách),

- jestli pacienta trápí *bolest na hrudi* bylo zaznamenáno v 93,7 % (v 1049 zprávách),
- otázky na *palpitaci* byly zaznamenány v 94,2 % (v 1054 zprávách),
- odpovědi na atribut *otoky* byly nalezeny v 93,8 % (v 1050 zprávách),
- *výška* byla zaznamenána v 97,9 % (v 1096 zprávách),
- *diabetes mellitus* v 95,9 % zpráv (v 1073 zprávách) a tak dále.

Zde se ukazuje, že aplikace ADAMEK není úplně dobře navržena, protože správně by tato aplikace, pokud by lékař některý atribut nevyplnil, neměla postoupit k dalšímu kroku a měla by vyžadovat vyplnění všech atributů.

4. 5. Atributy Minimálního datového modelu pro kardiologii zakódované pomocí SNOMED CT a MKN-10

Disertační práce ukazuje, několik příkladů atributů z Minimálního datového modelu pro kardiologii, kterým bylo přiděleno ConceptID z klasifikačního systému SNOMED CT [48]. Prvním předpokladem kódování, je ale přeložení názvu atributů do anglického jazyka, jelikož v současné době existuje pouze americká, britská, španělská a německá verze.

Jelikož je Mezinárodní klasifikace nemocí jednou z mála mezinárodních medicínských klasifikací, které jsou přeložené do českého jazyka, pokusila jsem se zakódovat termíny Minimálního datového modelu pro kardiologii právě pomocí této klasifikace, což je v disertační práci zobrazeno v podrobné tabulce. Pro srovnání jsou zde uvedeny rovněž kódy atributů MDMK v systému SNOMED CT [49].

4. 6. Výpočet míry diverzity vybraných atributů a jejich kategorií v textových a strukturovaných lékařských zprávách

Tradiční míry diverzity

Pro kompletní charakteristiku atributu je nezbytné stanovit si kategorie každého atributu. Pro daný atribut určíme kategorie $A_1, \ldots, A_{k-1}$ a případné další možnosti shrneme do kategorie "další" a tuto kategorii označíme jako A_k. Označme u sledované populace pravděpodobnostní rozdělení těchto kategorií symbolem $\mathbf{p} = (p_1, \ldots, p_k)$, $\sum_{i=1}^{k} p_i = 1$.

Gini-Simpsonův index $H_{GS}(\mathbf{p})$ se vypočítá jako

$$H_{GS}(\mathbf{p}) = 1 - \sum_{i=1}^{k} p_i^2 . \qquad (1)$$

Gini-Simpsonův index má své hodnoty v intervalu $[0, (k - 1)/k]$, kde dolní hranice 0 je dosaženo pouze tehdy, pokud existuje pouze jedna kategorie studovaného atributu a horní hranice $(k - 1)/k$ pro $\mathbf{p} = \mathbf{u} = (1/k, 1/k, \ldots, 1/k)$ pro rovnoměrné rozdělení pravděpodobnosti. Původně byl tento index navržen Ginim [50] jako míra nerovnosti v příjmech a později diskutován Simpsonem [51] jako míra ekologické diverzity.

Shannonův informační index $H_S(\mathbf{p})$ se vypočítá jako

$$H_S(\mathbf{p}) = -\sum_{i=1}^{k} p_i \log p_i . \qquad (2)$$

Shannonův informační index má hodnoty v intervalu [0; log k], kde dolní hranice 0 je dosaženo pouze tehdy, jestliže existuje pouze jedna kategorie atributu a horní hranice log k pro rovnoměrné rozdělení pravděpodobnosti $\mathbf{p} = \mathbf{u} = (1/k, \ldots, 1/k)$.

Je těžké dát obecnou přednost jedné z těchto dvou měr. Někteří výzkumníci jsou více obeznámeni s Shannonovou entropií a je pro ně jednodušší interpretovat konkrétní číselné hodnoty $H_S(\mathbf{p})$ než ty $H_{GS}(\mathbf{p})$. Na druhou stranu, Gini-Simpsonův index je tradiční mírou diverzity.

Kromě výše jmenovaných tradičních měr diverzity uvádí práce [52] následující míry diverzity: Havrdova a Charvátova entropie řádu α, párová Shannonova entropie, Rényiho entropie řádu α a γ-entropie. Bylo ukázáno, že některé z nich jsou speciálními případy obecnějšího konceptu f-diverzity [63], [64].

f-diverzita a relativní f-diverzita

Shannonova informace $I_S(X; Y)$ je v teorii informace definována jako míra asociace mezi dvěma atributy X a Y**.**

$$I_S(X;Y) = \sum_{x,z} p(x,y) \log \frac{p(x,y)}{p(x).p(y)}, \tag{3}$$

kde $p(x; y)$ jsou společné pravděpodobnosti a $p(x)$; $p(y)$ marginální pravděpodobnosti kategorií atributů X a Y.

Shannonova informace $I_S(X; Y)$ je nezáporná a rovna 0 pouze tehdy, jestliže jsou atributy nezávislé. Maximální informací je Shannova entropie, které je dosaženo, jestliže $Y = X$. V případě, že atribut X má kategorie $A_1, A_2, \ldots, A_k$, vyskytující se s pravděpodobnostmi $p_1, p_2, \ldots, p_k$, potom je *Shannonova entropie* atributu X stejná jako Shannonův informační index $H_S(\mathbf{p}) = -\sum_{i=1}^{k} p_i \log p_i$. Dále budeme nazývat tuto míru diveryity *Shannonovou diverzitou*.

Shannonova informace může být zobecněna na f-informaci

$$I_f(X;Y) = \sum_{x,y} f\left(\frac{p(x,y)}{p(x)\cdot p(y)}\right) p(x).p(y), \tag{4}$$

kde $f(t)$ je konvexní funkcí intervalu $[0;\infty)$, ryze konvexní v $t = 1$ s $f(1) = 0$. Více detailů o f-informaci odvozené z konceptu f-divergence najdete v práci Vajdy [53]. V případě $f(t) = t \log t$ se f-informace $I_f(X; Y)$ redukuje na Shannonovu informaci $I_S(X; Y)$, která je široce využívaná při rozpoznávání obrazců a při podpoře rozhodování, viz např. [54-57]. f-informace byla poprvé systematicky studována Zvárovou [58], která odvodila maximální hodnotu f-informace a nazvala ji f-entropií. V případě, že X je atribut s kategoriemi $A_1, A_2, \ldots, A_k$ a rozdělením pravděpodobnosti $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)$, potom je *f-entropie* atributu X

$$H_f(\mathbf{p}) = \sum_{i=1}^{k} p_i^2 f(1/p_i) + f(0) \sum_{i=1}^{k} p_i (1 - p_i). \tag{5}$$

f-entropie $H_f(\mathbf{p})$ může být interpretována jako průměrná nepředvídatelnost individuálních kategorií A_i atributu X [59]. V tomto smyslu je f-entropie $H_f(\boldsymbol{p})$ mírou diverzity závislé na distribuci $\mathbf{p}$. $H_f(\boldsymbol{p})$ se bude nazývat *f-diverzitou*, jestliže bude navíc splňovat následující podmínky:

- $H_f(\mathbf{p})$ je nezáporné,
- $H_f(\mathbf{p})$ dosahuje minimální hodnoty v případě, že jedna kategorie se vyskytuje s pravděpodobností 1,
- $H_f(\mathbf{p})$ dosahuje maximální hodnoty v případě, že $\mathbf{p} = \mathbf{u}$ je rovnoměrné rozložení,
- $H_f(\mathbf{p})$ je symetrická funkce $\boldsymbol{p}$,
- $H_f(\mathbf{p})$ je konkávní funkce systému všech pravděpodobnostních rozdělení $\mathbf{p}$.

f-diverzita byla poprvé zavedena Zvárovou [60] a detailněji diskutována v [61]. Vidíme, že $H_f(\mathbf{p})$ je součen dvou výrazů, kde druhý výraz je dobře známý Gini-Simpsonův index $H_{GS}(\mathbf{p})$ vynásobený konstantou $f(0)$. Dále budeme nazývat Gini-Simpsonův index *Gini-Simpsonovou diverzitou*. V článku [62] bylo dokázáno, že f-diverzity mohou být nalezeny mezi f-entropiemi splňujícími podmínku $g(t) = (f(t) - f(0))/t$ je konkávní funkcí. Potom f-entropie $H_f(\mathbf{p})$ atributu X bude dosahovat své maximální hodnoty pro rovnoměrné rozložení kategorií $\boldsymbol{p} = \mathbf{u}$. Vidíme, že Gini-Simpsonova diverzita $H_{GS}(\mathbf{p})$ je f-diverzita s $f(t) = t - 1$ pro $t > 1$, jinak $f(t) = 0$. Podobně, Shannonova diverzita je f-diverzita s $f(t) = t \log t$.

Relativní f-diverzita $RH_f(\mathbf{p})$ byla definována v [60] jako f-diverzita $H_f(\mathbf{p})$ dělena f-diverzitou rovnoměrného rozložení $H_f(\mathbf{u})$ jako

$$RH_f(\mathbf{p}) = H_f(\mathbf{p}) / H_f(\mathbf{u})$$

a byl zde uveden příklad uvádějící hodnoty Gini-Simpsonovy diverzity, diverzity řádu α, Shannonovy a párové Shannonovy diverzity.

Míry vzácnosti, vlastní a marginální f-diverzita

V případě, že X je atribut s kategoriemi $A_1, \ldots, A_k$ a pravděpodobnostním rozdělením $\mathbf{p} = (p_1, \ldots, p_k)$, potom podle Patila a Tailiee [62] závisí vzácnost kategorie A_i pouze na číselné hodnotě p_i. Označením vzácnosti kategorie A_i jako $R(p_i)$ je *index diverzity* spojený s mírou vzácnosti R jeho průměrnou vzácností počítanou jako

$$\sum_{i=1}^{k} p_i R(p_i). \tag{6}$$

Tři široce používané indexy diverzity jsou:
Počet kategorií (Diverzita počtu kategorií)

$$H_{NA} = k - 1 \quad \text{s } R(p_i) = (1 - p_i) = (1 - p_i)/p_i, \tag{7}$$

Gini-Simpsonův index (Gini-Simpsonova diverzita)

$$H_{GS}(\mathbf{p}) = \sum_{i=1}^{k} p_i(1 - p_i) \text{ s } R(p_i) = 1 - p_i \tag{8}$$

a Shannonův index (Shannonova diverzita)

$$H_S(\mathbf{p}) = -\sum_{i=1}^{k} p_i \log p_i \text{ s } R(p_i) = -\log p_i. \tag{9}$$

Tyto tři indexy diverzity patří do rodiny indexů diverzity typu β [62], definované jako

$$R_i(p_i) = \begin{cases} \left(1 - p_i^{\beta}\right)/\beta & if\ \beta \geq -1, \beta \neq 0. \\ -\log p_i & if\ \beta = 0. \end{cases} \tag{10}$$

Vidíme, že pro β = 0 získáme Shannovou diverzitu, pro β = 1 Gini-Simpsonovu diverzitu a pro β = -1 Diverzitu počtu kategorií. Jak bylo ukázáno výše, všechny tyto indexy diverzity patří do rodiny f-diverzit.

Představme si nyní koncept vlastní f-diverzity [63], což je zobecnění vzácnosti představené Patilem a Tailiem [62]. *Vlastní f-diverzita j*-té kategorie je definována jako

$$R_{f,j}(\mathbf{p}) = p_j f\left(1/p_j\right) + f(0)\left(1 - p_j\right). \tag{11}$$

Potom lze dokázat, že f-diverzitu spočítáme z vlastních f-diverzit jako

$$\begin{aligned} H_f(\mathbf{p}) &= \sum_{i=1}^{k} p_i \left(p_i f\left(1/p_i\right) + f(0)\left(1 - p_i\right)\right) \\ &= \sum_{i=1}^{k} p_i R_{f,i}(\mathbf{p}). \end{aligned} \tag{12}$$

f-diverzita $H_f(\mathbf{p})$ je tedy váženým průměrem vlastních f-diverzit $R_{f,i}(\boldsymbol{p})$.

Pro často používanou Shannovu diverzitu je *Shannonova vlastní diverzita* rovna

$$R_{S,j}(\mathbf{p}) = -\log\left(p_j\right) \tag{13}$$

známá v teorii informace jako *vlastní informace*. Podobně pro Gini-Simpsonovu diverzitu je *Gini-Simpsonova vlastní diverzita* rovna

$$R_{GS,j}(\mathbf{p}) = 1 - p_j. \tag{14}$$

Jiný pohled na význam *j*-té kategorie dostáváme, jestliže nebudeme rozlišovat mezi jinými kategoriemi. V tomto případě pracujeme formálně se dvěmi kategoriemi (dichotomie) s pravděpodobnostmi p_j a $1 - p_j$. *Marginální f-diverzita j*-té kategorie je definována jako

$$H_{f,j}(\mathbf{p}) = p_j^2 f\left(1/p_j\right) + \left(1 - p_j\right) f\left(1/\left(1 - p_j\right)\right) + 2 f(0) p_j \left(1 - p_j\right). \tag{15}$$

Dále uveďme relativní vlastní diverzitu a relativní marginální diverzitu [63], [64]. Definujme *relativní vlastní diverzitu j*-té kategorie jako

$$RR_{f,j}(\mathbf{p}) = R_{f,j}(\mathbf{p})/H_f(\mathbf{p}) \tag{16}$$

a *relativní marginální diverzitu j*-té kategorie definujeme jako

$$RH_{f,j}(\mathbf{p}) = H_{f,j}(\mathbf{p})/H_f(\tilde{\mathbf{u}}), \tag{17}$$

kde $\tilde{\mathbf{u}}$ = (½, ½).

Diverzita vybraných atributů a jejich kategorií v textových a strukturovaných lékařských zprávách

Analyzovali jsme 110 textových lékařských zpráv a 1119 strukturovaných lékařských zpráv z Městské nemocnice v Čáslavi. V disertační práci jsou shrnuty výsledky analýzy stejných vybraných atributů sebraných v textových a strukturovaných lékařských zprávách. Kategorizace těchto atributů byla provedena podle MDMK v 1119 strukturovaných zprávách a kategorie byly tvořeny jako hodnoty atributů zaznamenaných ve volném textu 110 textových lékařských zpráv. Jak již bylo zmíněno výše, pro vybrané atributy v textových zprávách jsme našli počet kategorií a vypočítali jsme Diverzitu počtu kategorií.

Každá textová lékařská zpráva byla čtena a analyzována jednotlivě, jedna po druhé, a všechny možné druhy zapisování vybraných atributů MDMK byly zvýrazněny a zaznamenány. Jelikož bylo nalezeno daleko více způsobů, jak byly vybrané znaky zapisovány v textových zprávách než ve strukturovaných zprávách, můžeme vidět, že Diverzita počtu kategorií je mnohem vyšší u textových zpráv než u strukturovaných zpráv.

Transformovali jsme kategorie atributů z textových zpráv tak, že jsme každé kategorii textových zpráv přiřadili jí nejbližší kategorii z MDMK. Odhadli jsme pravděpodobnosti kategorií MDMK pro všechny atributy z textových a strukturovaných lékařských zpráv bez chybějících pozorování a vypočítali jsme Gini-Simpsonovy diverzity, Gini-Simpsonovy vlastní diverzity a relativní marginální diverzity. Jak můžeme vidět z (14), Gini-Simpsonova vlastní diverzita kategorie je vyjádřena jako pravděpodobnost její komplementární kategorie. Proto s klesající pravděpodobností kategorie se zvyšuje Gini-Simpsonova diverzita. Souhrn všech Gini-Simpsonových vlastních diverzit pro daný atribut je roven počtu jeho kategorií minus jedna. Gini-Simpsonova relativní marginální diverzita kategorie se zvyšuje, když pravděpodobnost kategorie se přibližuje k ½. V případě, že vybraný atribut má pouze dvě kategorie, potom mají Gini-Simpsonovy relativní marginální diverzity těchto dvou kategorií stejnou hodnotu.

V textových zprávách hodně hodnot chybělo. Předpokládali jsme, že se jedná o negativní nálezy, které nebyly zaznamenány. Důvodem mohlo být například to, že lékař z předchozích atributů poznal, že sledovaný atribut již nemůže být přítomen, a proto ho nezaznamenal. Můžeme vidět, že kromě Gini-Simpsonovy relativní diverzity pro atribut "Alergie", jsou všechny vypočítané Gini-Simpsonovy relativní diverzity ve strukturovaných lékařských zprávách významně menší na 5% hladině ($p<0,05$), než v textových lékařských zprávách. Rozdíl pro atribut Alergie není na hladině 5 % významný. Byly použity statistické testy využívající Z statistiku se standardizovaným normálním rozdělením založeným na odhadech Gini-Simpsonovy diverzity. Nicméně, předpokládáme, že odhady vypočítané z textových lékařských zpráv budou ovlivněny velkým množstvím chybějících pozorování.

5. DISKUSE

Svojí prací jsem se snažila o ověření praktické použitelnosti mezinárodně používaných terminologických slovníků, tezaurů, ontologií a klasifikací a to konkrétně tak, že jsem studovala atributy Minimálního datového modelu pro kardiologii, které jsem dohledávala v mezinárodních klasifikačních systémech. Při této práci jsem využívala UMLS Metatezaurus.

Při mapování jsem čelila několika problémům – nejednoznačnosti při mapování a nemožnosti provést mapování z důvodu neexistenci odpovídajícího termínu v klasifikačních systémech. Velkým problémem při využití nomenklatur a metatezaurů ve zdravotnictví v České republice zůstává neexistence českých terminologických systémů či jejich vhodných českých překladů.

Atributy MDMK jsem primárně dohledávala v klasifikaci SNOMED CT, jelikož SNOMED CT je používán v HL7 verze 3.

Při mapování atributů MDMK na český překlad Mezinárodní klasifikace nemocí, vyplynulo, jak už samotný název klasifikace napovídá, že je možné tuto klasifikaci použít zejména pro zakódování nemocí, syndromů, patologických stavů, poranění, obtíží a jiných důvodů pro styk se zdravotnickými službami, tj. toho typu informací, které bývají registrovány lékařem. Bohužel, pomocí této klasifikace tedy nemůžeme zakódovat řadu atributů Minimálního datového modelu pro kardiologii, jako např. rodinný stav, vzdělání, psychickou zátěž, fyzickou zátěž, tělesnou aktivitu, kouření, pití alkoholu, fyzikální vyšetření (hmotnost, výška, tělesná teplota, obvod pasů, obvod boků, BMI, WHR, atd.), laboratorní vyšetření (celkový cholesterol, HDL-cholesterol) a ani popis EKG. MNK se hodí pouze pro části Minimálního datového modelu pro kardiologii týkající se osobní anamnézy a pro současné potíže možného kardiovaskulárního původu.

Při lexikální analýze textových zpráv, jsem zjistila, že při zapisování výsledků vyšetření pomocí volného textu zůstává plno znaků nezaznamenáno. K tomu může docházet z několika důvodů. Lékaři nemají přesně danou osnovu, podle které by měli postupovat a může se stát, že na některé znaky mohou zapomenout. Dalším důvodem, proč nejsou některé znaky v textové zprávě zaznamenány, může být fakt, že lékařům ze znalosti předchozích znaků vyplyne, že další znak nemůže být přítomen a proto se již na něj dále nezeptají a nezaznamenají ho. Z textové zprávy ale nevyplyne, zda skutečně byly u pacienta zjišťovány tyto základní informace, z jejichž hodnot lékaři hodnoty dalších znaků sami svými znalostmi vyvodili.

Vzhledem k velké diverzitě zaznamenávání atributů v textových zprávách, jsou tyto zprávy daleko složitější pro počítačové zpracování než strukturované zprávy a mnoho informací z textových zpráv může být v tomto procesu ztraceno a proto je velice důležité vytvořit standardizovanou terminologii lékařských pojmů, která bude pro všechny lékaře závazná.

6. ZÁVĚR

Hlavním cílem disertační práce byla informačně lexikální analýza českých lékařských zpráv a využitelnost mezinárodních klasifikačních systémů v českém zdravotnickém prostředí. Byl proveden detailní rozbor stávajících mezinárodních klasifikačních systémů a podrobně rozebrán Unified Medical Language System, sloužící jako konverzní nástroj pro velkou část těchto mezinárodních klasifikací. Disertační práce také popisuje Minimální datový model pro kardiologii, což je minimální soubor znaků, které je potřeba sledovat u pacientů z hlediska aterosklerotického kardiovaskulárního onemocnění, aby mohl být pacient následně zařazen mezi osoby nemocné či rizikové. Na jeho základě byla vytvořena softwarová Aplikace Datového Modelu EuroMISE centra – Kardio (ADAMEK), která slouží ke sběru dat v ambulanci preventivní kardiologie EuroMISE centra, která je spravována Městskou nemocnicí Čáslav.

Disertační práce popisuje analýzu lékařských zpráv, která byla založena na atributech Minimálního datového modelu pro kardiologii. Byly použity lékařské zprávy psané volným textem a strukturované lékařské zprávy uložené v softwarové aplikaci ADAMEK. Práce je zaměřena na jazyk česky psaných lékařských zpráv a na aplikaci výše zmíněných mezinárodních klasifikačních systémů v MDMK.

Analýzou textových lékařských zpráv bylo zjištěno, že zapisování pomocí volného textu je velice nesourodé a nestandardizované. Rozdíly najdeme nejenom ve zprávách od různých lékařů, ale i jednotliví lékaři často zapisují stejné koncepty v různých tvarech. Největšími problémy pro další počítačové zpracování jsou překlepy, různá délka

zkracovaných výrazů a používání synonym, které vede k velké nejednoznačnosti a dokonce i k nepřesnostem a nedorozuměním. Jelikož v dnešní době dochází k velkému rozvoji elektronických zdravotních záznamů, je nezbytné vytvořit standardizovanou terminologii. Při naší práci jsme došli k závěrům, že standardizovaná terminologie by přinesla výhody jak lékařům, tak i pacientům, administrátorům, softwarovým vývojářům a plátcům. Standardizovaná klinická terminologie by pomohla poskytovatelům lékařské péče tak, že by jim poskytla snáze dostupné a kompletní informace, které náleží k procesu zdravotnické péče (chorobopis pacienta, nemoci, léčby, laboratorní výsledky, atd.) a to by vedlo k lepším výsledkům v péči o zdraví.

Pro sémantickou interoperabilitu je nezbytné využívání mezinárodních klasifikačních systémů. Současné zdravotnické informační systémy umožňují sbírat různé klinické informace, tyto systémy jsou propojeny s klinickými znalostními databázemi, mohou vyhledávat data, shromažďovat data, analyzovat data, vyměňovat si data a mají i plno dalších funkcí. Jako nejlepším klasifikačním systémem se zatím jeví SNOMED CT, který může poskytnout základy pro tyto funkce. Informační systémy mohou využít koncepty, hierarchie a vztahy jako společný referenční bod. Tato terminologie může, například, usnadnit podporu rozhodování, statistické zpracovávání, sledování veřejného zdraví, zdravotnický výzkum a analýzy nákladů. Z tohoto důvodu byly pro moji práci využity zejména klasifikační systémy SNOMED CT a MKN-10, jako jeden z mála mezinárodních klasifikačních systémů přeložených do českého jazyka, pomocí nichž byly namapovány atributy Minimálního datového modelu pro kardiologii.

I přes problémy, které při využití mezinárodních nomenklatur a metatezaurů ve zdravotnictví v České republice přetrvávají, jako je jejich neexistence v českém jazyce nebo příliš velká nebo naopak malá granularita atributů, je jejich využití prvním a nezbytným krokem k umožnění sémantické interoperability heterogenních systémů zdravotních záznamů. Dostatečná sémantická interoperabilita těchto systémů je totiž základem pro sdílenou zdravotní péči, která vede k efektivitě ve zdravotnictví, finančním úsporám i snížení zátěže pacientů, a proto se ve své práci snažím analyzovat, jak mezinárodních klasifikačních systémů využít co nejlépe pro potřeby českého zdravotnictví.

Tato disertační práce také ukazuje novou aplikaci pro měření diverzity lékařských zpráv psaných v jakémkoli jazyce. Tato nová aplikace je založena na obecných konceptech diverzit a byla odvozena z f-diverzity, relativní f-diverzity, vlastní f-diverzity a marginální f-diverzity [63], [64]. Byly porovnány diverzity vybraných atributů v textových lékařských zprávách a porovnány výsledky s diverzitami stejných atributů ve strukturovaných lékařských zprávách. Ukázalo se, že atributy v textových lékařských zprávách mají větší diverzitu než stejné atributy ve strukturovaných lékařských zprávách. V obou typech zpráv byly opět použity atributy z Minimálního datového modelu pro kardiologii. Až na atribut „Alergie“ u Gini-Simpsonovy relativní diverzity, všechny vypočítané Gini-Simpsonovy diverzity ve strukturovaných lékařských zprávách byly menší než v textových lékařských zprávách. Ukázalo se, že využitím vybraných měr diverzity můžeme porovnávat neurčitost v zaznamenávání informací ve strukturovaných a textových lékařských zprávách.

Efektivní péče o zdraví vyžaduje dobré informace. Bezpečná a vhodná výměna klinických informací je nezbytná k zajištění kontinuity péče o pacienty a to v různých časech, na různých místech a u různých poskytovatelů zdravotní péče.

LITERATURA

[1] Bleich H. L., Slack W. V.: Reflections on electronic medical record: When doctor will use them and when they will not, Int. J. Med. Inform. 2010; 79: 1-4.

[2] Hoerbst A., Kohl C. D., Knaup P., Ammenwerth E.: Attitudes and behaviors related to the introduction of electronic health records among Austrian and German citizens, Int. J. Med. Inform. 2010; 79: 81-89.

[3] Rinner C., Janzek-Hawlat S., Sibinovic S., Duftschmid G.: Semantic Validation of Standard-based Electronic Health Record Documents with W3C XML Schema. Method Inf Med 2010; 49, preprint online

[4] Oemig F., Blobel B.: Semantic Interoperability Adheres to Proper Models and Code Systems: A Detailed Examination of Different Approaches for Score Systems. Methods Inf Med 2010; 49 (2): 148-155

[5] Lopez D.M., Blobel B.: A development framework for semantic interoperable health information systems. Int J Med Inform 2009; 78 (2): 83-103

[6] Garde S., Knaup P., Hovenga E.J.S., Heard S.: Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for openEHR Archetypes. Methods Inf Med 2007; 46 (3): 332-343

[7] Nagy M., Hanzlíček P., Přečková P., Kolesa P., Mišúr J., Dioszegi M., Zvárová J.: Building Semantically Interoperable EHR Systems Using International Nomenclatures and Enterprise Programming Technique. In eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge. (Eds. Blobel, B.; Pharow, P.; Zvárová, J.; Lopez, D.) Amsterdam: IOS Press, 2008: 105-110

[8] Nagy M., Hanzlíček P., Přečková P., Říha A., Dioszegi M., Seidl L., Zvárová J.: Semantic Interoperability in Czech Healthcare Environment Supported by HL7 version 3. Methods Inf Med 2010; 49 (2): 186-195

[9] Zvárová J., Hanzlíček P., Nagy M., Přečková P., Zvára K., Seidl L., Bureš V., Šubrt D., Dostálová T., Seydlová M.: Biomedical Informatics Research for Individualized Life-long Shared Healthcare. In: Biocybernetics and Biomedical Engineering, 2009; 29 (2): 31-41

[10] Přečková P., Špidlen J., Zvárová J.: Usage of the International Nomenclatures and Metathesauruses in Shared Healthcare in the Czech Republic. Acta Informatica Medica, 2005; (13): 201-205

[11] Přečková P., Zvárová J., Špidlen J.: International Nomenclatures in Shared Healthcare in the Czech Republic. In: Proceedings of 6th Nordic Conference on eHealth and Telemedicie „From Tools to Services“ (Ed.: Doupi P.), 2006, 45-46

[12] Elkin P. L., Trusko B. E., Koppel R., Speroff T., Mohrer D., Sakji S., Gurewitz I., Tuttle M., Brown S. H.: Secondary Use of Clinical Data. In Seamless Care – Safe Care. (Eds. Blobel B., Hvannberg E., Gunnarsdóttir), IOS Press, 2010, 14-29

[13] Mezinárodní statistická klasifikace nemocí a přidružených zdravotních problémů. Desátá revize. Instruktážní příručka. ÚZIS ČR. Aktualizovaná druhá verze k 1. 1. 2010.

[14] Ústav zdravotnických informací a statistiky České republiky: Mezinárodní statistická klasifikace nemocí a přidružených zdravotních problémů (MKN-10). ©WHO, ©ÚZIS ČR. domovská stránka na internetu dostupná z http://www.uzis.cz/cz/mkn/index.html (citováno 31. 5. 2011).

[15] World Health Organizstion: International Classification of Diseases (ICD). ©2011, domovská stánka na internetu, dostupná z http://www.who.int/classifications/icd/en/ (citováno 31. 5. 2011).

[16] Stausberg J., Lehmann N., Kaczmarek D., Stein M.: Reability of diagnose coding with ICD-10. International Journal of Medical Informatics 2008; 77: 50-57
[17] The International Health Terminology Standards Development Organisation: SNOMED Clinical Terms®, domovská stránka na internetu, dostupná z http://www.ihtsdo.org/snomed-ct/ (citováno 31. 5. 2011).
[18] The International Health Terminology Standards Development Organisation: SNOMED Clinical Terms® User Guide. ©2002-2009, July 2009 International Release, 1-70.
[19] Schulz S., Hanser S., Hahn U., Rodgers J.: The Semantics Procedures and Diseases in SNOMED® CT. Methods Inf Med 2006; 45: 354-8
[20] Cornet R.: Definitions and Qualifiers in SNOMED CT. Methods Inf Med 2009; 48: 177-183
[21] U. S. National Library of Medicine, National Institutes of Health: Medical Subject Headings. Domovská stránka na internetu dostupná z http://www.nlm.nih.gov/mesh/ (citováno 31. 5. 2011).
[22] Gault Lora V., Schultz M.: Variations in Medical Subject Headings (MeSH) mapping: from the natural language of patron terms to the controlled vocabulary of mapped lists. J Med Libr Assoc. 2002; 90(2): 173–180
[23] Regenstrief Institute, Inc.: Logical Observation Identifiers Names and Codes (LOINC®), ©1994-2011, domovská stránka na internetu dostupná z http://www.regenstrief.org/medinformatics/loinc/ (citováno 31. 5. 2011).
[24] Khan A. N., Griffith S. P., Moore C., Russell D., Rosario A. C., Jr., Bertolli J.: Standardizing Laboratory Data by Mapping to LOINC. J Am Med Inform Assoc. 2006; 13(3): 353–355
[25] World Health Organization: International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3). © 2011, domovská stránka na internetu, dostupná z http://www.who.int/classifications/icd/adaptations/oncology/en/ (citováno 31. 5. 2011)
[26] Louis D. N., Ohgaki H., Wiestler O. D., Cavenee W. K., Burger P.C., Jouvet A., Scheithauer B.W.,Kleihues P.: The 2007 WHO Classification of Tumours of the Central Nervous System. Acta Neuropathol. 2007; 114(2): 97–109
[27] TNM Classification Help: Manual for Cancer Staging. domovská stánka na internetu dostupná z http://cancerstaging.blogspot.com/ (citováno 31. 5. 2011).
[28] Brierley J.: The evolving TNM cancer staging system: an essential component of cancer care. CMAJ. 2006; 174(2): 155–156
[29] American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders (DSM). ©2011, domovská stránka na internetu dostupná z http://psych.org/MainMenu/Research/DSMIV.aspx (citováno 31. 5. 2011)
[30] U.S. National Library of Medicine, National Institute of Health: Unified Medical Language System® (UMLS®): domovská stránka na internetu, dostupná z http://www.nlm.nih.gov/pubs/factsheets/umls.html (citováno 31. 5. 2011)
[31] Han S-B., Choi J.: The comparative study on concept representation between the UMLS and the clinical terms in Korean Medical Records. International Journal of Medical Informatics 2005; 74: 67-76
[32] Campbell J.R., Olivek D.E., Shortliffe: UMLS: towards a collaborative approcah for solving terminologic probels, J.Am. Med. Inform. Assoc. 1998; 5: 12-16
[33] Adášková J., Anger Z., Aschermann M., Bencko V., Berka P., Filipovský J., Goláň L., Grus T., Grünfeldová H., Haas T., Hanuš P., Hanzlíček P., Holcátová I., Hrach K., Jiroušek R., Kejřová E., Kocmanová D., Kolář J., Kotásek P., Králíková E., Krupařová M., Kyloušková M., Malý M., Mareš R., Matoulek M., Mazura I., Mrázek V.,

Novotný L., Novotný Z., Pecen L., Peleška J., Prázný M., Pudil P., Rameš J., Rauch J., Reissigová J., Rosolová H., Rousková B., Říha A., Sedlak P., Slámová A., Somol P., Svačina Š, Svátek V., Šabík D., Šimek S., Škvor J., Špidlen J., Štochl J., Tomečková M., Umnerová V., Zvára K., Zvárová J.: Návrh minimálního datového modelu pro kardiologii a softwarová aplikace ADAMEK. Interní výzkumná zpráva EuroMISE Centra – Kardio. Praha, říjen 2002.

[34] Tomečková M.: Minimální datový model kardiologického pacienta – výběr dat. Cor et Vasa, 2002; 44 (4), Suppl.: 123

[35] Mareš R., Tomečková M., Peleška J., Hanzlíček P., Zvárová J.: Uživatelská rozhraní pacientských databázových systémů – ukázka aplikace určené pro sběr dat v rámci Minimálního datového modelu kardiologického pacienta. Cor et Vasa, 2002; 44 (4), Suppl.: 76

[36] Health Level Seven International: HL7 Version 3 Standards, ©2007-2011, domovská stránka na internetu, dostupná z http://www.hl7.org/ (citováno 31. 5. 2011).

[37] Jung B., Grimson J.: Synapses/SynEx goes XML. Studies in Health Technology and Informatics, 1999; 68: 906-911.

[38] Centre for Health Informatics and Multiprofessional Education (CHIME): The Good European Health Record, dostupný z: http://www.chime.ucl.ac.uk/work-areas/ehrs/GEHR/ (citováno 31. 5. 2011).

[39] ©Ocean Informatics 2008, domovská stránka na internetu, dostupná z http://www.oceaninformatics.com/ (citování 31. 5. 2011).

[40] Lipka J., Mukenšnábl Z., Horáček F., Bureš V.: Současný komunikační standard českého zdravotnictví DASTA. In: Zvárová J., Přečková P. (eds.): Informační technologie v péči o zdraví, EuroMISE s.r.o., Praha, 2004: 52-59.

[41] Ministerstvo zdravotnictví České republiky: Datový standard MZ ČR, Národní číselník laboratorních položek MZ Č R a Národní zdravotnický informační systém, domovská stránka na internetu dostupná z http://ciselniky.dasta.mzcr.cz/ (citováno 31. 5. 2011).

[42] Eryiğit G., Nivre J., Oflazer K.: Dependency Parsing of Turkish. Computational Linguistics. 2008; 34(3): 357-389.

[43] Přečková P.: Jazyk lékařských zpráv. Doktorandský den 2007. MATFYZPRESS 2007: 75-79

[44] Zvára K, Kašpar V. Identification of Units and Other Terms in Czech Medical Records. European Journal for Biomedical Informatics 2010; 6 (1): 78-82.

[45] Ringlestetter C., Schulz K. U., Mihov S.: Orthographic Errors in Web Pages: Toward Cleaner Web Corpora. Computational Linguistics. 2006, 32(3): 295-340

[46] Mareš R., Tomečková M., Peleška J., Hanzlíček P., Zvárová J.: Uživatelská rozhraní pacientských databázových systémů – ukázka aplikace určené pro sběr dat v rámci Minimálního datového modelu kardiologického pacienta. Cor et Vasa, 2002; 44 (4) Suppl.:76.

[47] Mareš R.: ADAMEK – uživatelská příručka. EuroMISE centrum – Kardio. 2002.

[48] Přečková P.: SNOMED CT a jeho využití v Minimálním datovém modelu pro kardiologii. Doktorandský den 2008. MATFYZPRESS 2008: 99-105

[49] Přečková P.: Mezinárodní klasifikace nemocí a její využití v Minimálním datovém modelu pro kardiologii. In: Doktorandský den 09. (Ed.: Hakl F.) - Praha, MATFYZPRESS 2009: 97-101.

[50] Gini C: Variabilità e Mutabilità. Studi Economico-Giuridici della R. Univ. di Cagliari. 3, 1912; Part 2 80.

[51] Simpson EH. Measurement of diversity. Nature 1949; 163: 688.

[52] Chakraborty R, Rao CR. Measurement of genetic variation for evolutionary studies. Statistical and Medical Sciences. Elsevier Science Publ, 1991; 271-316.
[53] Vajda I. Theory of Statistical Inference and Information. Boston, Kluwer, 1989.
[54] Zvarova J, Studeny M. Information theoretical approach to constitution and reduction of medical data. Int J Med Inf 1997; 45: 65-74.
[55] Peng H, Long F, Ding Ch. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Mas-Relevance and Min-Redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005; 27 (8): 1226-1238.
[56] Benish WA. Intuitive and Axiomatic Arguments for Quantifying Diagnostic Test Performance in Units of Information. Methods Inf Med 2009; 48: 552-557.
[57] Blokh D, Zurgil N, Stambler I, Afrimzon E, Shafran Y, Korech E, Sandbank J, Deutsch M. An Information-theoretical Model for Breast Cancer Detection. Methods Inf Med 2008; 47: 322-557.
[58] Zvárová J. On measures of statistical dependence. Časopis pro pěstování matematiky 1974; 99: 15-29.
[59] Zvárová J, Vajda I. On genetic information, diversity and distance. Methods Inf Med 2006; 2: 173-179.
[60] Zvárová J. Information Measures of Stochastic Dependence and Diversity: Theory and Medical Informatics Applications. Disertační práce, Akademie věd ČR, 1998.
[61] Zvárová J, Mazura I. Stochastická genetika. Karlova Univerzita, Karolinum, Praha, 2001.
[62] Patil GP, Tailie C. Diversity as a concept and its measurement. Journal of American Statistical Association, 1982; 77: 548-561.
[63] Zvárová J, Zvára K. Stochastic modelling of biodiversity: f-diversity, self f-diversity and marginal f-diversity. In: J. Hrebicek and J. Holcik eds. Proceedings of the 6th Summer School on Computational Biology, Deterministic and Stochastic Modelling in Biology and Medicine, Akademické nakladatelství CERM, Brno 2010, 108-119.
[64] Přečková P., Zvárová J., Zvára K.: Measuring Diversity of Medical Reports by Categorized Attributes and International Classification Systems. BMC Medical Informatics and Decision Making 2011, submitted.

Seznam publikací doktoranda v tomto uspořádání:

1. publikace *in extenso,* které jsou podkladem disertace

a) s impact factorem (uvést hodnotu IF)

1. Nagy M., Hanzlíček P., **Přečková P.,** Říha A., Dioszegi M., Seidl L., Zvárová J.: *Semantic Interoperability in Czech Healthcare Environment Supported by HL7 Version 3.* Methods Inf Med 2010; 49 (2): 186-195, **IF 1,69**
2. **Přečková P.,** Zvárová J., Zvára K.: *Measuring Diversity of Medical Reports by Categorized Attributes and International Classification Systems*. BMC Medical Informatics and Decision Making 2011, submitted. **IF 1,9**

b) bez IF

1. **Přečková P.**: *Digitální knihovny, biomedicínská data a znalosti.* In: Doktorandský den ´04. . (Ed.: Hakl F.), Praha. MATFYZPRESS 2004: 78-84
2. **Přečková P.**: *SPECIALIST lexikon a čeština.* In: Informační technologie v péči o zdraví. (Ed.: Zvárová J., Přečková P.), Praha, EuroMISE 2004: 124-127
3. **Přečková P.,** Špidlen J., Zvárová J.: *Užití mezinárodních nomenklatur ve sdílené zdravotnické péči v ČR.* In: Sdílení informací o zdraví. (Ed.: Zvárová J., Přečková P.), Praha, EuroMISE 2005: 60-63
4. **Přečková P.**, Špidlen J., Zvárová J.: *Usage of the International Nomenclatures and Metathesauruses in Shared Healthcare in the Czech Republic.* Acta Informatica Medica 2005 (13): 201-205
5. **Přečková P.:** *Mezinárodní nomenklatury a metatezaury ve zdravotnictví.* In: Doktorandský den 05. (Ed.: Hakl F.) - Praha, MATFYZPRESS 2005: 109-116
6. Kolesa P., **Přečková P.**: *Effective Creation of Czech Biomedical Ontologies.* In: Integrating Biomedical Information: From eCell to ePatient. (Ed.: Reichert A., Mihalas G., Stoicu-Tividar L., Schulz S., Engelbrecht R.) - Amsterdam, Aka 2006: 305-310
7. Kolesa P., **Přečková P.**: *Tools for Czech Biomedical Ontologie Creation.* In: Ubiquity: Technologies for Better Health in Aging Societies. (Ed.: Hasman A., Haux R., van der Lei J., De Clercq E., Roger France F. H.). IOS Press 2006: 775-780
8. **Přečková P.,** Zvárová J., Špidlen J.: *International Nomenclatures in Shared Healthcare in the Czech Republic.* In: Proceedings of 6th Nordic Conference on eHealth and Telemedicie „From Tools to Services" (Ed.: Doupi P.) 2007: 45-46
9. Hanzlíček P., **Přečková P.**, Zvárová J.: *Semantic Interoperability in the Structured Electronic Health Record.* Ercim News, 2007, 69: 52-53
10. **Přečková P.**: *Jazyk lékařských zpráv*. In: Doktorandský den 07. (Ed.: Hakl F.), Praha, MATFYZPRESS 2007: 75-79
11. Nagy M., Hanzlíček P., **Přečková P.**, Kolesa P., Mišúr J., Dioszegi M., Zvárová J.: *Building Semantically Interoperable EHR Systems Using International Nomenclatures and Enterprise Programming Technique.* In eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge. Amsterdam : (Eds. Blobel, B.; Pharow, P.; Zvárová, J.; Lopez, D.) IOS Press, 2008: 105-110
12. **Přečková P.**: *SNOMED CT a jeho využití v Minimálním datovém modelu pro kardiologii*. In: Doktorandský den 08. (Ed.: Hakl F.), Praha, MATFYZPRESS 2008: 99-105

13. Nagy M., Hanzlíček P., Dioszegi M., Zvárová J., **Přečková P.**, Seidl L., Zvára K., Bureš V., Šubrt D.: *Applied Information Technologies for Development of Continuous Shared Health Care.* In: Cesnet Coference 2008. Security, Middleware, and Virtualization - glue of Future Networks. Prague, Czech Republic. Cesnet, z. s. p. o. (Eds. Krčmářová G., Sojka P.), 2008: 131-138
14. Nagy M., Hanzlíček P., Dioszegi M., Zvárová J., **Přečková P.**, Seidl L., Zvára K., Bureš V., Šubrt D.: *Realizace elektronického zdravotního záznamu pro sdílenou péči s využitím mezinárodních standardů a nomenklatur.* In: DATAKON 2008. Brno: Masarykova Univerzita (Eds. Řepa, V.; Svatoš, O.), 2008: 197-205
15. Zvárová J., Dostálová T., Hanzlíček P., Nagy M., Seydlová M., Hippmann R., **Přečková P.**, Červená I., Psutka J., Šmídl L., Zvára jr., K., Seidl L., Eliášová H., Šimková H.: *Voice-supported Electronic Health Record in Dentistry.* In: INFOLAC 2008 - AAIM. Buenos Aires : Asociación Argentina de Informática Médica. 2008: 1-3
16. **Přečková P.**: *SNOMED CT and its use in the Minimal Data Model for Cardiology*. In: 5th Meeting of the Doctoral Schools of the Charles University (Prague) and Louis Pasteur University (Strasbourg). Strasbourg : ULP, 2008
17. Nagy M., Hanzlíček P., Dioszegi M., Zvárová J., **Přečková P.**, Seidl L. , Zvára K., Bureš V., Šubrt D.: *Electronic Health Record for Shared Care Based on International Standards and Nomenclatures in Czech National Environment.* TeleMed & eHealth 2008. Optimizing Patient Centred Care - The Role of eHealth. London: The Royal Society of Medicine.
18. Zvárová J., Hanzlíček P., Nagy M., **Přečková P.**, Zvára K., Seidl L., Bureš V., Šubrt D., Dostálová T., Seydlová M.: *Biomedical Informatics Research for Individualized Life-long Shared Healthcare.* In: Biocybernetics and Biomedical Engineering, 2009 (9), 2: 31-41
19. **Přečková P.**: *Mezinárodní klasifikace nemocí a její využití v Minimálním datovém modelu pro kardiologii*. In: Doktorandský den 09. (Ed.: Hakl F.), Praha, MATFYZPRESS 2009: 97-101
20. **Přečková P.,** Zvárová J.: *Language of Czech Medical Reports*. ISCB 2009. Prague, 2009:145
21. **Přečková P.,** Zvárová J.: *The Role of International Nomenclatures and. Standards in Travel Shared Health Care.* Travel Health Informatics and Telehealth. Istanbul : EFMI, 2009 - (Mihalas, G.; Saka, O.; Mazzoleni, T.; Blobel, B.; Pharow, P.; Gülkesen, K.) 2009:162-169. (European Notes in Medical Informatics. 5-1).
22. Daněk J., Hliníková P., **Přečková P.**, Dostálová T., Nedoma J., Nagy M.: *Modelling of the Temporomandibular Joints and the Role of Medical Informatics in Stomatology*. ICCSA 2010 (Eds.: D. Taniar et al.), Part IV, LNCS 6016, 2010: 62-71
23. Zvárová J., Lhotská L., Přibík V., Adášková J., Brechlerová D., Hanzlíček P., Kopecký M., Papíková V., Potůček J., **Přečková P.,** Říha A., Svátek V., Šárek M., Zitová B., Zvára K.: *Biomedicínská data a znalosti*. Univerzita Karlova v Praze – Nakladatelství Karolinum. 2010
24. Nagy M., **Přečková P.**, Seidl L., Zvárová J.: *Challenges of Interoperability Using HL7 v3 in Czech Healthcare*. EFMI STC 2010 Reykjavík Iceland, Stud Health Technik Inform. 2010 (155):122-8.
25. **Přečková P.:** *Jazyk českých lékařských zpráv a klasifikační systémy v medicíně*. In: Sémantická interoperabilita v biomedicíně a zdravotnictví. (Svačina Š, Zvárová J. eds.). EuroMISE s.r.o 2010: 58-65.
26. **Přečková P.**: *Language of Czech Medical Reports and Classification Systems in Medicine*. European Journal for Biomedical Informatics 2010; 6 (1): 58-65